Universidad Politécnica de Madrid
Escuela Técnica Superior de Ingenieros de Telecomunicación

Contribución a los modelos de estimación de la calidad percibida en servicios de vídeo sobre Internet mediante parámetros objetivos

Tesis Doctoral

Joaquín Navarro Salmerón
Ingeniero de Telecomunicación

Año 2015

Departamento de Ingeniería de Sistemas Telemáticos
Escuela Técnica Superior de Ingenieros de Telecomunicación

Tesis Doctoral

Contribución a los modelos de estimación de la calidad percibida en servicios de vídeo sobre Internet mediante parámetros objetivos

Autor

Joaquín Navarro Salmerón

Ingeniero de Telecomunicación

Director

Francisco González Vidal

Doctor Ingeniero de Telecomunicación

Año 2015

Tribunal nombrado por el Magfco. y Excmo. Sr. Rector de la Universidad Politécnica de Madrid, el día ..... de ..... de 2015.

Presidente:.....

Vocal:.....

Vocal:.....

Vocal:.....

Secretario:.....

Suplente:.....

Suplente:.....

Realizado el acto de defensa y lectura de Tesis el día ..... de ..... de 2015 en la E.T.S. de Ingenieros de Telecomunicación de Madrid.

Calificación:.....

EL PRESIDENTE LOS VOCALES

EL SECRETARIO

"We learn something every day, and lots of times it's that what we learned the day before was wrong."

— Bill Vaughan

Resumen

En los últimos años el consumo de servicios de vídeo se ha incrementado de forma notable y se espera que dicha tendencia continúe en los próximos años. Los servicios de streaming de vídeo Over-The-Top (OTT), en los que se centra esta tesis, constituyen uno de los principales motores de dicho crecimiento. A diferencia de los servicios Internet Protocol Television (IPTV), que utilizan una red controlada en la que se pueden implementar mecanismos de Quality of Service (QoS), los servicios de streaming de vídeo OTT se prestan sobre Internet, por lo que llevan asociados interesantes desafíos desde un punto de vista técnico. Uno de los mayores desafíos técnicos a los que se enfrentan los servicios de streaming de vídeo OTT es mantener un nivel de Quality of Experience (QoE) que satisfaga a sus usuarios, por lo que es necesario contar con técnicas y herramientas que permitan monitorizar la calidad percibida por los usuarios de estos servicios.

El streaming de vídeo OTT supone un cambio de filosofía en comparación con otras técnicas de streaming más tradicionales como RTP/RTSP. Los servicios de vídeo OTT suelen seguir el paradigma Dynamic Adaptive Streaming over HTTP (DASH), que se basa en sustituir los servidores de streaming tradicionales por servidores web que ponen a disposición de los clientes los contenidos de vídeo codificados en varias versiones con distinto nivel de calidad. Cada una de estas versiones o representaciones está dividida en pequeños fragmentos o segmentos que los clientes pueden solicitar mediante el protocolo HTTP. Los clientes pueden solicitar diferentes niveles de calidad en función de los parámetros que consideren más adecuados (ancho de banda de la red, resolución de pantalla, tipo de códec, etc.), lo que les permite adaptarse a condiciones cambiantes del entorno. Como se puede ver, el paradigma DASH ha trasladado el control de la sesión del servidor al cliente y ha sustituido los servidores de streaming por servidores web que simplemente sirven los segmentos de vídeo que los clientes solicitan. Además se esta simplificación de los servidores de streaming, existen otras ventajas asociadas a DASH, como son la utilización de Content Delivery Network (CDN), la compatibilidad con NATs y firewalls, etc.

En esta tesis doctoral se lleva a cabo la propuesta de un conjunto de modelos cuyo objetivo es estimar la calidad percibida por los usuarios de los servicios de vídeo

basados en DASH. Más concretamente, partiendo de la definición del servicio como un conjunto de componentes de servicio, se desarrollan modelos parciales que estiman la calidad percibida asociada a cada uno de estos componentes: calidad de vídeo, calidad de audio, degradaciones asociadas a la transmisión, etc. Cada una de estas estimaciones de calidad percibida se combinan en un modelo global que estima la calidad percibida total del servicio.

Palabras clave: calidad percibida, calidad de experiencia, QoE, vídeo, streaming, DASH, MPEG-DASH.

Abstract

In recent years video services consumption has increased notably and it is expected that this trend will continue in the foreseeable future. This thesis focuses on Over-the-Top (OTT) video services, which are one of the main drivers of the aforementioned growth. In contrast to IPTV services, which are provided over managed IP networks where Quality of Service (QoS) mechanisms can be implemented, OTT video streaming services are deployed on the Internet. This entails some technological challenges, such as providing the service with enough Quality of Service (QoE) to please the users. In this context, techniques and tools that allow the perceived quality to be monitored are needed.

OTT video streaming entails a paradigm shift in comparison with more traditional streaming techniques such as RTP/RTSP. Most of OTT video services are based on Dynamic Adaptive Streaming over HTTP (DASH). That means that traditional streaming servers are replaced by web servers that make available to the users video content coded in several versions (representations) with different quality. Each of these versions is divided into small fragments (segments) that users can request using the HTTP. The user client has to decide which quality level it requests, taking into account different parameters such as network bandwidth, user equipment capabilities, etc. This allows user clients to dynamically adapt to changing conditions. As can be seen, DASH transfers the session control from server to client, replacing streaming servers with web servers that handle video segments requests. As there is no need to maintain a session state in the server, HTTP streaming is highly scalable. In addition to that, HTTP streaming has other benefits. First, HTTP server technology has become a commodity, so HTTP streaming is a cost effective technology. Furthermore, as the Internet has evolved to efficiently support HTTP, Content Delivery Networks (CDNs) can be used to reduce long-haul traffic, HTTP outgoing connections can traverse firewalls, etc.

This thesis proposes a set of models to estimate the perceived quality of DASH video streaming services. In first place, the service is formally defined as a set of service components. Then, partial models are developed to estimate the perceived quality of each of those components: video quality, audio quality, network degradations, etc. Finally, all these perceived quality estimations are combined using a global model to

estimate the total perceived quality of the service.

Keywords: perceived quality, quality of experience, QoE, video, streaming, DASH, MPEG-DASH.

Agradecimientos

Si hoy estoy escribiendo estas líneas es gracias al apoyo y a la ayuda de muchas personas que me han acompañado a lo largo de esta etapa de mi vida.

Quiero dar las gracias en primer lugar a Enri, por estar ahí cada día, haciendo que todo sea más fácil simplemente por estar a su lado. Agradecer también a mis padres y hermanos el apoyo incondicional y el cariño que cada día me transmiten aunque estén lejos.

Agradecer también a todo el grupo RSTI la confianza que han puesto en mí a lo largo de esta etapa. A los profesores Julio, Enrique, Víctor, Manolo y como no, a mi director de tesis, Paco, que me ha guiado en momentos de dudas y cuyo optimismo ha contribuido enormemente a alcanzar esta meta. Sin su ayuda no habría sido posible llegar hasta aquí.

Dar las gracias también a mis compañeros de laboratorio. A los antiguos compañeros, Vicente, Pedro y Alberto. A los actuales compañeros Pilar, Verónica y sobre todo a Carlos y a Mario, que me acompañan en el día a día, compartiendo este camino y haciendo que el ambiente de trabajo sea excepcional.

Índice general

Resumen ix

Abstract xi

Agradecimientos xiii

Índice de figuras xxi

Índice de tablas xxv

1. Introducción **1**
1.1. Contexto y motivación 1
1.2. Objetivos 5
1.3. Estructura de la memoria 7

2. Marco conceptual **9**
2.1. Concepto general de calidad 9
2.2. Calidad de servicio 11
2.2.1. Definiciones 11
2.2.2. Parámetros de rendimiento 12
2.2.3. Clases de tráfico y clases de servicio 14
2.2.4. Mecanismos de implementación de QoS en redes IP 16
2.3. Calidad percibida 17
2.3.1. Definiciones 17
2.3.2. Modelos generales de calidad percibida 19
2.3.3. Medida de QoE en servicios de telecomunicaciones 23
2.4. MPEG-DASH 26
2.4.1. Introducción 26
2.4.2. Streaming adaptativo 27
2.4.3. Arquitectura de referencia y alcance del estándar 28
2.4.4. Estructura del fichero MPD 29

2.4.5. Formato de los segmentos 30
2.5. Codificación de vídeo 30
2.5.1. Introducción 30
2.5.2. Evolución de los estándares de codificación de vídeo 31
2.5.3. Proceso de codificación 33
2.6. Resumen y conclusiones 35

3. Estimación de la calidad percibida en servicios de streaming multimedia sobre Internet 37
3.1. Introducción 37
3.2. Planteamiento general del modelo 37
3.2.1. Escalas de calidad y nomenclatura 39
3.3. Modelo global de estimación de QoE de un servicio de streaming de vídeo a partir de las valoraciones de calidad de sus componentes 44
3.3.1. Componentes continuos 45
3.3.2. Componentes puntuales 47
3.4. Componentes continuos 49
3.4.1. Estimación del factor de calidad audiovisual para flujos sincronizados 49
3.4.2. Sincronización audio-vídeo 60
3.4.3. Degradación de calidad debida a la transmisión 62
3.5. Componentes puntuales 63
3.5.1. Cambio de canal 63
3.5.2. Acceso aleatorio 68
3.6. Resumen y conclusiones 73

4. Modelo de estimación de calidad de vídeo 75
4.1. Introducción y motivación 75
4.2. Revisión del estado del arte 76
4.2.1. Proyectos Video Quality Expert Group 78
4.2.2. Recomendaciones International Telecommunication Union (ITU) 81
4.2.3. Artículos científicos 89
4.2.4. Conclusiones extraídas del estado del arte 99
4.3. Desarrollo del modelo 100
4.3.1. Selección del modelo de referencia 101
4.3.2. Selección de la base de datos de secuencias de vídeo de prueba . 101
4.3.3. Medidas de VQM-VFD 104
4.3.4. Entrenamiento del modelo 109
4.3.5. Evaluación del modelo 120

4.4. Resumen y conclusiones 125

5. Modelo de degradación de calidad debida a la transmisión 127
5.1. Introducción 127
5.2. Revisión del estado del arte 127
5.2.1. Buffering inicial y eventos de rebuffering 128
5.2.2. Adaptación del nivel de calidad 135
5.3. Desarrollo del modelo 141
5.3.1. Introducción 141
5.3.2. Metodología: experimentos de evaluación subjetiva de calidad de vídeo 142
5.3.3. Tiempo de buffering inicial 144
5.3.4. Eventos de rebuffering 146
5.3.5. Adaptación de calidad de vídeo 150
5.4. Análisis de la influencia de la red en las variables del modelo 155
5.4.1. Aproximación al problema de manera analítica 156
5.4.2. Aproximación al problema mediante simulación de red 158
5.5. Resumen y conclusiones 178

6. Conclusiones y líneas de trabajo futuras 181
6.1. Análisis de los objetivos 181
6.1.1. Propuesta de un modelo global de estimación calidad percibida para servicios de streaming de vídeo adaptativo OTT 181
6.1.2. Propuesta de un modelo de estimación de calidad percibida de vídeo 184
6.1.3. Propuesta de un modelo de estimación de degradación en la calidad percibida asociada a la red y a los mecanismos de transmisión 185
6.2. Difusión de resultados 187
6.3. Líneas de trabajo futuro 187

A. Modelo de descripción de servicios 189
A.1. Introducción y motivación 189
A.2. Marco de referencia 190
A.3. Descripción del modelo 192
A.3.1. Objetivos 192
A.3.2. Elementos del modelo 192
A.3.3. Representación gráfica 195
A.4. Metodología para la aplicación del modelo de descripción de servicios al dominio de los servicios multimedia 196

A.4.1. Descripción de la metodología 196

B. Secuencias de vídeo utilizadas **203**
B.1. Modelo de calidad de vídeo 203
B.2. Degradación asociada al tiempo de buffering inicial 206
B.3. Degradación asociada al tiempo de rebuffering 206
B.4. Degradación asociada al número de eventos de rebuffering 207
B.5. Degradación asociada a los mecanismos de adaptación de calidad 207

C. Plataforma web de evaluación subjetiva de calidad de vídeo **209**
C.1. Introducción 209
C.2. QualityCrowd2 210
C.3. Modificaciones realizadas a QualityCrowd2 210
C.3.1. Sustitución del reproductor de vídeo 210
C.3.2. Simulación de eventos de buffering inicial y rebuffering 211
C.3.3. Extensión de la sintaxis QC-script 212

D. Comparativa y selección de herramientas de simulación de redes **213**
D.1. OPNET Modeler 213
D.2. NS-2 214
D.3. NS-3 214
D.4. OMNeT++ 215
D.5. NetSim 215
D.6. Selección de la herramienta de simulación 216

Bibliografía **217**

Acrónimos **235**

Índice de figuras

2.1. Modelo de QoE de Oliver 20
2.2. Modelo de calidad de Hardy 23
2.3. Arquitectura genérica de MPEG-DASH 28
2.4. Estructura del MPD de MPEG-DASH 29
2.5. Tipos de tramas MPEG 34

3.1. Modelo de referencia del servicio de streaming multimedia sobre Internet 38
3.2. Relación entre escala R y escala MOS según ITU-T G.107 41
3.3. Relación propuesta entre escala MOS y escala R 43
3.4. Función $f(Q_C)$ propuesta 48
3.5. Arquitectura de un modelo de calidad multimedia según ITU-T J.148 . 50
3.6. Calidad audiovisual en función de la calidad de los flujos de audio y vídeo [Garcia and Raake, 2009] 55
3.7. Factor de degradación de calidad asociado al lipsync 61
3.8. Calidad del cambio de canal con varianza nula 66
3.9. Degradación asociada al tiempo de cambio de canal con varianza nula . 69
3.10. Degradación de calidad asociada al error en el acceso aleatorio 73

4.1. Proceso de cálculo de VQM. [International Telecommunication Union (ITU), 2004c] 82
4.2. Proceso de cálculo de VQuadHD. [ITU, 2011b] 88
4.3. VQM_VFD para las secuencias de vídeo VQEG-HD1 104
4.4. VQM_VFD para las secuencias de vídeo VQEG-HD2 105
4.5. VQM_VFD para las secuencias de vídeo VQEG-HD3 105
4.6. VQM_VFD para las secuencias de vídeo VQEG-HD5 106
4.7. VQM_VFD para las secuencias de vídeo VQEG-CommonSet 106
4.8. VQM_VFD para la secuencia de vídeo VQEG-HD3SRC4 109
4.9. Valores del parámetro a en función de ASI y ATI 110
4.10. Valores del parámetro b en función de ASI y ATI 111
4.11. Red neuronal: perceptrón multicapa 112

4.12. Arquitectura de la red neuronal utilizada en el modelo 113
4.13. Bias y overfitting 114
4.14. Curva de aprendizaje 115
4.15. Rendimiento de la red neuronal entrenada con Levenberg-Marquardt . . 120
4.16. MSE de la red neuronal entrenada con Levenberg-Marquardt (curva de aprendizaje) 121
4.17. Estimación de VQM_VFD para secuencias de prueba no utilizadas en el entrenamiento (Levenberg-Marquardt) 122
4.18. Rendimiento de la red neuronal entrenada con regularización bayesiana 124
4.19. Estimación de VQM_VFD para secuencias de prueba no utilizadas en el entrenamiento (regularización bayesiana) 125

5.1. Calidad en función del tiempo de rebuffering. [Tan et al., 2006] 128
5.2. Calidad en función del número de eventos de rebuffering. [Tan et al., 2006] 128
5.3. Tasa de abandono en función del tiempo de buffering inicial para diferentes duraciones de vídeo. [Krishnan and Sitaraman, 2012] 131
5.4. Tasa de abandono en función del tiempo de buffering inicial para distintas tecnologías de red de acceso. [Krishnan and Sitaraman, 2012] 132
5.5. Tiempo de reproducción en función del tiempo de rebuffering. [Krishnan and Sitaraman, 2012] 133
5.6. Trayectoria de adaptación óptima para distintos tipos de contenido. [Cranley et al., 2006] 136
5.7. Modelo de estimación de trayectoria de adaptación óptima. [Cranley et al., 2007] 137
5.8. Involucramiento en función de la frecuencia de cambios de calidad. [Balachandran et al., 2012] 139
5.9. Metodología seguida en el desarrollo del modelo de degradación debida a la transmisión 142
5.10. Efecto del tiempo de buffering inicial: valoraciones subjetivas y modelo propuesto 145
5.11. Efecto del tiempo de buffering inicial: comparativa con otros modelos . . 146
5.12. Efecto del tiempo de rebuffering: valoraciones subjetivas y modelo propuesto 147
5.13. Efecto del tiempo de rebuffering: comparativa con otros modelos 148
5.14. Efecto del número de eventos de rebuffering con respecto al tiempo total de rebuffering 150
5.15. Efecto del número de eventos de rebuffering: valoraciones subjetivas y modelo propuesto 151
5.16. Efecto del número de eventos de rebuffering: comparativa con otros modelos 152

5.17. Diagrama de clases del modelo de simulación de streaming de vídeo adaptativo 159
5.18. Ejemplo de trazas de simulación de streaming de vídeo adaptativo . . . 161
5.19. Topología de la red simulada 163
5.20. Comportamiento del algoritmo de adaptación simplificado 168
5.21. Traza del nivel de calidad solicitado por un usuario con canal D_i=20Mbps 176

A.1. Diagrama UML del modelo de descripción de servicios 195
A.2. Descripción del servicio de televisión lineal utilizando el modelo propuesto 198
A.3. Componente de servicio "Visualización de video": bloques arquitecturales e implementaciones para un sistema de vídeo OTT 200
A.4. Descripción del servicio de Video on Demand (VoD) utilizando el modelo propuesto 202

Índice de tablas

2.1. Clases de tráfico según 3rd Generation Partnership Project (3GPP) . . . 14
2.2. Clases de tráfico según ITU 15
2.3. Clases de servicio según ITU 15
2.4. Clases de servicio según ITU: aplicaciones e implementación 16

3.1. Parámetros de ajuste para la función de conversión entre escala R y escala MOS 43
3.2. Parámetros de ajuste para la función $f(Q_C)$ 49
3.3. Coeficientes del modelo ITU-T G.1070 53
3.4. Coeficientes del modelo de García, versión 2009 55
3.5. Tipos de contenido contemplados en el modelo de García, versión 2011 . 56
3.6. Coeficientes del modelo basado en componentes de calidad de García et al, versión 2011 56
3.7. Coeficientes del modelo basado en factores de degradación de García et al, versión 2011 57
3.8. Coeficientes del modelo ITU-T P.1201.2 57
3.9. Coeficientes del modelo de calidad audiovisual propuesto (adaptación de ITU-T P.1201.2) 58
3.10. Parámetros de ajuste del modelo de vídeo ITU-T P.1201.2 59
3.11. Parámetros de ajuste del modelo de audio ITU-T P.1201.2 60
3.12. Efecto de la sincronización audio-vídeo en función del contenido 62
3.13. Umbrales aproximados de aceptabilidad y detección del lipsync en función del tipo de contenido 62

4.1. Parámetros de ajuste del modelo Joskowicz et al 93
4.2. Bases de datos de secuencias de vídeo de prueba HD 102
4.3. Parámetros de ajuste VQM_VFD para las secuencias VQEGHD 107
4.4. Técnicas de reducción de bias y overfitting 116
4.5. MSE para secuencias de prueba no utilizadas en el entrenamiento (Levenberg-Marquardt). Tasa de bit de 1 a 12 Mbps 123

4.6. MSE para secuencias de prueba no utilizadas en el entrenamiento (Levenberg-Marquardt). Tasa de bit de 2 a 12 Mbps 123
4.7. Comparativa de algoritmos de entrenamiento en términos de MSE para secuencias de prueba no utilizadas en el entrenamiento. Tasa de bit de 1 a 12 Mbps 124
4.8. Comparativa de algoritmos de entrenamiento en términos de MSE para secuencias de prueba no utilizadas en el entrenamiento. Tasa de bit de 2 a 12 Mbps 126

5.1. Niveles de degradación de QoE del modelo [Mok et al., 2011] 129
5.2. Parámetros de ajuste del modelo de degradación asociada al tiempo de buffering inicial 144
5.3. Parámetros de ajuste del modelo de degradación asociada al tiempo de rebuffering 147
5.4. Parámetros de ajuste del modelo de degradación asociada al número de eventos de rebuffering 150
5.5. Resultados del experimento de evaluación de calidad en escenarios de adaptación del nivel de calidad 153
5.6. Líneas de banda ancha fijas por segmento y velocidad [CNMC, 2012] . . 162
5.7. Velocidades consideradas en los canales D_i 163
5.8. Capacidades de los canales para el experimento de simulación 1 164
5.9. Resultados agregados del experimento de simulación 1 165
5.10. Resultados de la simulación 1.1 166
5.11. Resultados de la simulación 1.2 166
5.12. Resultados de la simulación 1.3 167
5.13. Resultados de la simulación 1.4 167
5.14. Resultados de la simulación 1.5 167
5.15. Resultados de la simulación 1.6 167
5.16. Resultados agregados del experimento de simulación 2 170
5.17. Resultados de la simulación 2.1 170
5.18. Resultados de la simulación 2.2 170
5.19. Resultados de la simulación 2.3 170
5.20. Resultados de la simulación 2.4 170
5.21. Resultados de la simulación 2.5 171
5.22. Resultados de la simulación 2.6 171
5.23. Resultados agregados del experimento de simulación 3 173
5.24. Resultados de la simulación 3.1 174
5.25. Resultados de la simulación 3.2 174
5.26. Resultados de la simulación 3.3 174

5.27. Resultados de la simulación 3.4 174
5.28. Resultados de la simulación 3.5 174
5.29. Resultados de la simulación 3.6 175

B.1. Secuencias de vídeo VQEGHD1 203
B.2. Secuencias de vídeo VQEGHD2 204
B.3. Secuencias de vídeo VQEGHD3 204
B.4. Secuencias de vídeo VQEGHD5 205
B.5. Secuencias de vídeo VQEGHDCommonSet 205
B.6. Secuencias de vídeo utilizadas en el experimento de evaluación de calidad del tiempo de buffering inicial 206
B.7. Secuencias de vídeo utilizadas en el experimento de evaluación de calidad del tiempo de rebuffering 206
B.8. Secuencias de vídeo utilizadas en el experimento de evaluación de calidad del número de eventos de rebuffering 207
B.9. Secuencias de vídeo utilizadas en el experimento de evaluación de calidad de la adaptación de vídeo (1 de 2) 207
B.10. Secuencias de vídeo utilizadas en el experimento de evaluación de calidad de la adaptación de vídeo (2 de 2) 208

Capítulo 1

Introducción

1.1. Contexto y motivación

En los últimos cinco años el tráfico IP se ha quintuplicado y se espera que se triplique en los siguientes cinco años. Este incremento es posible gracias a la mejora de las redes de comunicación de banda ancha y viene impulsado fundamentalmente por el aumento del tráfico asociado a servicios de vídeo. En 2013 el tráfico de vídeo supuso el 66 % del total del tráfico, mientras que se espera que éste represente el 79 % del tráfico IP en 2018 [Cisco, 2014].

Estos datos engloban tanto vídeo transmitido a través de Internet, conocido como vídeo OTT, como vídeo transmitido mediante redes IP gestionadas por los operadores, IPTV. Si se comparan los datos de ambos sistemas se puede ver que el vídeo OTT ya en 2013 genera más tráfico que IPTV. Además, las predicciones auguran un gran aumento del tráfico de vídeo OTT (Compound annual growth rate (CAGR) del 30 %) mientras que la evolución del tráfico IPTV se cree que alcanzará una CAGR del 16 %.

Como se ha comentado, el vídeo OTT engloba a los sistemas de distribución de vídeo a través de Internet. Algunos de los proveedores de vídeo OTT más destacados actualmente son Netflix, Hulu, Amazon Prime Instant Video, etc. Por otro lado, los portales de vídeo de contenido generado por usuarios, como Youtube o Vimeo, contribuyen enormemente al total del tráfico de vídeo OTT. Tecnológicamente hablando, los servicios de vídeo OTT se basan en el paradigma del streaming adaptativo sobre HTTP. Esta tecnología consiste en codificar los contenidos en diversas versiones con distinto nivel de calidad, dividir el vídeo en pequeños fragmentos (segmentos) y ponerlos a disposición de los usuarios a través de un servidor web. Así pues, es el cliente el que gestiona la sesión de streaming, solicitando en cada momento el nivel de calidad que considere adecuado en función de las condiciones de la red, de la capacidad del dispositivo utilizado, etc. Esta tecnología reduce los costes en el lado del servidor si se compara con otras tecnologías de streaming en la que la sesión está controlada por el

servidor, ya que cualquier servidor web puede hacer las veces de servidor de streaming. Por otro lado, la utilización de HTTP permite que este tipo de sistemas se beneficien de cachés y CDN de manera directa.

El aumento en el tráfico IP de vídeo, viene de la mano de una serie de factores fundamentales: el aumento de la penetración de los servicios de vídeo no tradicionales (vídeo OTT y IPTV), el aumento de los minutos de vídeo consumido y el aumento de la calidad de vídeo que los usuarios demandan.

En cuanto a los hábitos de consumo de vídeo y televisión, según el informe [PwC, 2013], realizado en Los Ángeles en 2013, los sistemas de suscripción tradicionales (satélite y cable) siguen siendo los productos dominantes del mercado. Sin embargo, es destacable que la mayoría de los participantes en la encuesta cuentan también con otras suscripciones adicionales (de tipo vídeo OTT, como Netflix, Hulu o Amazon Prime) las cuales satisfacen ciertas necesidades que el cable y el satélite no contemplan, como por ejemplo: tener la posibilidad de acceder al contenido en cualquier momento y en cualquier lugar, recomendaciones de contenido basadas en los contenidos consumidos anteriormente, etc. Según un estudio realizado por Park Associates, un 55 % de los hogares con banda ancha en Estados Unidos están suscritos a un servicio de vídeo OTT, siendo Netflix el servicio más contratado por los norteamericanos.

El contenido y la originalidad del mismo es otro factor determinante a la hora de elegir un tipo de suscripción. Para un 63 % del total de los encuestados la originalidad del contenido es un factor importante. Más aún para el segmento de encuestados de entre 25 y 34 años, ya que el porcentaje de personas que consideran importante la originalidad del contenido supone un 72 %.

Un informe elaborado por LRG afirma que en 2013 el 29 % de los usuarios de Netflix consumieron contenidos diariamente y un 70 % semanalmente, mientras que en 2010 estos porcentajes eran del 10 % y del 43 % respectivamente. Así pues, el número de programas consumidos mensualmente por usuarios de Netflix pasó de 9,9 en 2011 a 19,6 en 2013 [LRG, 2013].

Además de analizar los datos de Netflix, es interesante analizar también el aumento de la penetración del servicio Amazon Instant Video (Amazon Prime) en Estados Unidos. Este servicio, nacido en 2006 con el nombre de Amazon Unbox se ha conseguido extender al 20 % de hogares con conexión de banda ancha en Estados Unidos, doblando su tasa de penetración con respecto al año anterior. Por su parte, Hulu consiguió llevar su servicio al 12 % de los hogares norteamericanos con conexión de banda ancha [M2M, 2014].

Las cifras del mercado de los servicios de vídeo OTT en Europa no son tan elevadas como en Estados Unidos. Sin embargo, algunos de los principales distribuidores de contenidos (como Netflix) están empezando a establecerse en Europa. Según Global-

Connect, en junio de 2014 la penetración de Netflix en Dinamarca era del 29 %, lo cual supone 725000 suscriptores. Otras cifras establecen un número de usuarios de Netflix en Noruega y Suecia de 380000 y 880000, lo cual supone una penetración del 20 % y del 17 % respectivamente.

Además, muchas cadenas de televisión (tanto gratuitas como de pago) están ofreciendo servicios de vídeo OTT que permiten a los usuarios ver tanto contenido en vivo como contenido en diferido. Por ejemplo, en Reino Unido la mitad de usuarios con acceso a Internet de banda ancha utilizan servicios de vídeo OTT, impulsados en parte por el éxito de iPlayer de la BBC.

Por otro lado, algunos operadores de red europeos se están posicionando como agregadores de contenidos OTT y se están involucrando cada vez más en el mercado audiovisual. Un ejemplo interesante es el caso de Jazztel, que canceló su servicio de vídeo basado en IPTV llamado Jazztelia y lo sustituyó por una solución OTT, Jazzbox, la cual permite acceder al servicio de Canal+ Yomvi. Otro ejemplo destacable es el caso de Telefónica, la cual se encuentra actualmente en pleno proceso de revisión por parte de la Comisión Nacional de los Mercados y de la Competencia de la compra de Canal+, lo cual supondría que Telefónica controlase aproximadamente el 60 % del mercado de la televisión de pago en España, además de importantes derechos sobre contenidos clave.

Teniendo en cuenta todo esto, las predicciones son esperanzadoras en cuanto al crecimiento de dicho mercado en Europa. Según Digital TV Research (DTR), en el año 2020, 59 millones de hogares europeos contarán con una suscripción a algún servicio de streaming OTT, siendo Netflix y Amazon los principales motores de este aumento en la penetración. Además, IHS predice que el mercado europeo representará el 20 % del total de suscriptores de Netflix a finales del 2015.

En España, el mercado de los servicios de streaming de vídeo OTT está bastante lejos del de otros países, tanto europeos como de otros continentes. En general, la penetración de los servicios de televisión de pago en España es relativamente baja. Según el último informe sobre "Consumos y gastos de los hogares españoles en los servicios de comunicaciones electrónicas; segundo semestre de 2013", ésta se cifra en un 21,2 %. Esta baja penetración se traduce por tanto en números relativamente bajos en cuanto a contrataciones de servicios de streaming OTT.

Los servicio líderes en televisión de pago en España son los siguientes:

- Canal+: 1,6 millones de suscriptores. Es el servicio de televisión de pago líder en España, el cual, como se comentó anteriormente, está en proceso de ser adquirido por Telefónica. Tecnológicamente se basa en un modelo de televisión por satélite DTH (Direct to Home).

- Movistar TV: conocido anteriormente como Imagenio, el servicio IPTV de Telefónica cuenta con 1,2 millones de suscriptores. El servicio incluye también acceso

OTT para los usuarios móviles.

- Ono: adquirido por Vodafone en marzo de 2014, la empresa de cable cuenta con casi 800000 suscriptores a TV de pago. Ono también ha lanzado un servicio multiscreen disponible en PC, Mac y iPad.

- Gol TV: servicio de TDT de pago con unos 237000 clientes.

Además de las iniciativas de vídeo OTT impulsadas por los operadores, generalmente para soportar multiscreen, existen otras compañías que ofrecen contenidos en streaming OTT en España: Wuaki.tv, Yomvi (Canal+), Filmin, Cineclick, Nubeox, TotalChannel y Magine. Estas empresas se basan fundamentalmente en dos modelos de negocio: el alquiler de contenidos y el modelo de suscripción premium con tarifa plana mensual. Además, algunos grupos mediáticos españoles ofrecen sus contenidos o parte de ellos de manera gratuita mediante una plataforma de vídeo OTT, como por ejemplo Mitele de Mediaset, AtresPlayer del grupo Atresmedia y el servicio de TV en directo y "a la carta" de RTVE.

En cuanto al número de usuarios de estas plataformas, en [Genbeta, 2014] se afirma que al comienzo de 2014 Wuaki contaba aproximadamente con un millón de usuarios (875000 en España y 125000 en el Reino Unido), país en el que la empresa catalana llegó tras ser adquirida por la japonesa Rakuten en 2012. Por otro lado, en [Hemerotek, 2014] se cifra el número de usuarios de Yomvi en casi 500000, lo cual supone aproximadamente el 29 % de los abonados al servicio de televisión por satélite de Canal+. Sin embargo, se debe destacar que del total de usuarios de Yomvi, solo 28000 lo son de manera independiente al servicio de televisión por satélite.

En cuanto a las cifras de los usuarios que utilizan los servicios catch-up de las cadenas de televisión, la plataforma Mitele de Mediaset es la más popular, con 3,7 millones de usuarios en 2014. El grupo RTVE consiguió atraer a 2,8 millones de usuarios, mientras que Atresmedia contó con 2,6 millones de usuarios [Ovum, 2014].

Como se desprende de los datos presentados, desde un punto de vista geográficamente global, en los últimos años se está produciendo un crecimiento considerable tanto de la penetración de los servicios de streaming de vídeo como del consumo que los usuarios hacen del mismo.

En cuanto a la calidad de los contenidos consumidos, ésta ha ido aumentando de forma progresiva a lo largo de los últimos años.

Por ejemplo, Netflix comenzó codificando el contenido con WMV3 a tasas de 500, 1000, 1600 y 2200 kbps para resoluciones de 720x480 píxeles. Después pasaron a utilizar VC1 Advanced Profile, que al ser más eficiente que WMV3 permitió reducir las tasas de bit a 375, 500, 1000 y 1500 kbps. En la siguiente iteración, empezaron a distribuir vídeo HD (720p) utilizando VC1AP a tasas de 2600 y 3800 kbps. Actualmente la

mayoría de contenidos que distribuye Netflix son vídeos Full HD 1080p, utilizando diversos codecs para poder adaptarse a la diversidad de dispositivos en los que el servicio puede consumirse. Desde mediados de 2014 Netflix está incluyendo en su catálogo algunos contenidos en calidad 4K Ultra High Definition Video (UHDV), con resolución 3840x2160 [Netflix, 2008], [Netflix, 2013] [Netflix, 2014].

La calidad de los contenidos emitidos por los distribuidores de contenido es un factor crucial en el negocio y un elemento diferenciador con respecto al resto de competidores del mercado. Por tanto, para estos actores es fundamental contar con herramientas o modelos que les permitan obtener información acerca de la calidad con la que se presta su servicio.

Existen dos puntos de vista desde los que abordar el estudio de la calidad. El primero de ellos, la calidad de servicio o QoS está orientada a la monitorización y al control de una serie de parámetros técnicos de rendimiento (anchos de banda, retardo, jitter, etc.). El segundo punto de vista, la calidad percibida o QoE, trata de evaluar y medir el nivel de satisfacción que el usuario percibe al consumir el servicio. Es evidente que ambos puntos de vista de la calidad están íntimamente relacionados. Sin embargo, mientras que la calidad de servicio es fácil de medir, la monitorización de la calidad percibida presenta todavía importantes retos.

Para las empresas distribuidoras de contenidos, como Netflix o Hulu, sería deseable contar con herramientas o modelos capaces de estimar la calidad que están percibiendo sus usuarios en tiempo real. Sin embargo, en la literatura no se han encontrado modelos globales que tengan en cuentan de manera unificada los distintos elementos o componentes que forman el servicio en su totalidad. Por otro lado, los nuevos mecanismos de distribución de vídeo, basados en HTTP y técnicas de adaptación de calidad en el lado del cliente, cambian el panorama en cuanto al tipo de degradaciones que los usuarios pueden percibir, con respecto a las técnicas de streaming clásicas basadas en RTP/UDP con control de sesión por parte del servidor. Esto hace que ciertos modelos parciales que se diseñaron para estimar la calidad en streaming basado en RTP no sean directamente aplicables a los nuevos escenarios de vídeo OTT.

Estas razones son las que motivan la investigación que se lleva a cabo en esta tesis, cuyos objetivos concretos se detallan en la siguiente sección.

1.2. Objetivos

El principal objetivo de esta tesis es desarrollar un modelo de estimación de calidad percibida para servicios de streaming de vídeo adaptativo sobre protocolos fiables y orientados a conexión, tomando como esquema de referencia el protocolo The Moving Picture Experts Group - Dynamic Adaptive Streaming over HTTP (MPEG-DASH).

Este modelo será un modelo sin referencia, es decir, no podrá tener acceso a la señal

audiovisual original (antes de ser transmitida por la red). Más concretamente, el modelo tendrá como datos de entrada un conjunto de parámetros objetivos y medibles desde el lado del cliente y deberá generar una valoración de calidad percibida en una escala numérica adecuada. Esto permite que la estimación de calidad pueda ser implementada en los dispositivos del cliente y pueda ser llevada a cabo en tiempo casi real.

Se debe destacar también que el modelo a desarrollar deberá ser un modelo global, es decir, deberá tener en cuenta no solo la calidad del vídeo recibido por el cliente, sino también la influencia que tienen el resto de componentes del servicio en la calidad percibida por el usuario.

En base a este objetivo global, se proponen los siguientes objetivos concretos:

- **Propuesta de un modelo global de estimación calidad percibida para servicios de streaming de vídeo adaptativo OTT**: este modelo deberá combinar las contribuciones de cada uno de los componentes del servicio a la calidad percibida por el usuario. Más concretamente, este modelo tendrá en cuenta la **calidad de vídeo, calidad de audio, degradación asociada a la sincronización entre el audio y vídeo, degradación asociada al efecto de la red y los mecanismos de transmisión, calidad asociada al tiempo de seeking (acceso aleatorio) y calidad asociada al tiempo de cambio de canal**. Combinando todos estos factores el modelo obtendrá una estimación de la calidad percibida en una escala numérica de 1 a 5, siendo 1 la calidad mínima y 5 la calidad máxima (escala MOS).

- **Propuesta de un modelo de estimación de calidad percibida de vídeo**: este modelo deberá ser capaz de estimar la calidad de vídeo, considerando las degradaciones introducidas en el proceso de **codificación** y sin utilizar la señal de vídeo original.

- **Propuesta de un modelo de estimación de degradación en la calidad percibida asociada a la red y a los mecanismos de transmisión**: este modelo deberá ser capaz de cuantificar la degradación en la calidad percibida que se puede producir como consecuencia de transmitir el flujo audiovisual a través de la red, utilizando mecanismos de streaming adaptativo sobre HTTP. En concreto, tendrá en cuenta el efecto del **tiempo de buffering inicial, el número de eventos de rebuffering, el tiempo total de dichos eventos de rebuffering y de los cambios en la calidad de vídeo**.

Para cada uno de estos objetivos se seguirá una metodología similar. En primer lugar se llevará a cabo un estudio de la literatura con el objetivo de identificar y analizar trabajos relacionados. Como resultado del análisis se deberá decidir si los

trabajos propuestos en la literatura se pueden aplicar a las necesidades concretas de la tesis. En caso negativo, se propondrán nuevos modelos que, partiendo del conocimiento adquirido en el estudio del estado del arte, satisfagan los requisitos de la tesis.

1.3. Estructura de la memoria

Esta tesis se organiza en los siguientes capítulos:

- Capítulo 1: Introduce el contexto, la motivación y los objetivos que persigue esta tesis doctoral.
- Capítulo 2: Conceptos generales sobre calidad, calidad de servicio, calidad percibida, el estándar MPEG-DASH y algunos conceptos de codificación de vídeo.
- Capítulo 3: Introduce el modelo global de estimación de calidad percibida en servicios de vídeo OTT. Describe los fundamentos de diseño del modelo y presenta cada uno de los componentes del mismo. Algunos de estos componentes, de especial relevancia, se tratan en capítulos independientes.
- Capítulo 4: Presenta el desarrollo del modelo de estimación de calidad de vídeo. Este modelo permite obtener una estimación de la métrica Video Quality Model for Variable Frame Delay (VQM_VFD) sin utilizar referencia, en contenidos de vídeo codificado en H.264, enfocado a resoluciones Full HD.
- Capítulo 5: En este capítulo se introducen modelos capaces de estimar la degradación que sufre la calidad percibida por los usuarios por el efecto de la red y de los mecanismos de transmisión utilizados. En concreto se estudia el efecto del tiempo de buffering inicial, tiempo de rebuffering, número de eventos de rebuffering y mecanismos de adaptación de calidad de vídeo.
- Capítulo 6: Conclusiones más relevantes del desarrollo de la tesis e introducción de líneas futuras de investigación.
- Anexo A: Definición de un modelo de descripción de servicios basado en componentes. Dicho modelo sirve como fundamento conceptual para el desarrollo del resto de modelos de estimación de calidad percibida, ya que permite describir un servicio complejo en función de un conjunto de componentes reutilizables.
- Anexo B: Incluye capturas de las secuencias de vídeo utilizadas en el desarrollo de los modelos de calidad.
- Anexo C: Describe la plataforma web de evaluación de calidad que se ha utilizado para obtener valoraciones de usuarios reales.

- Anexo D: Presenta una comparativa de herramientas de simulación de redes.

Es importante destacar que, debido a que en esta tesis se proponen distintos modelos de estimación de calidad (modelo de calidad global, modelo de calidad de vídeo, modelos de degradación asociados a la transmisión, etc.), **el análisis del estado del arte asociado a cada uno de ellos se ha realizado en su capítulo correspondiente**. Por esta razón, en el capítulo 2 de estado del arte no se describen propuestas concretas de modelos de estimación de calidad, sino conceptos más generales que abarcan todo el ámbito de la tesis.

Capítulo 2

Marco conceptual

En este capítulo se introducen una serie de conceptos generales que sirven para establecer el marco conceptual en el que se sitúa esta tesis. La revisión del estado del arte de cuestiones más específicas y técnicas se ha realizado dentro de cada capítulo del resto de la tesis, con el objetivo de facilitar la lectura y acercar lo máximo posible la descripción de los trabajos relacionados al punto concreto de la tesis donde son especialmente relevantes.

2.1. Concepto general de calidad

En esta sección se aborda el estudio del concepto general de calidad, sin ligar dicho estudio a ningún ámbito, tecnología o sistema concreto.

En la literatura se pueden encontrar diversas definiciones del concepto de calidad. A continuación se recogen algunas de las más destacadas:

- ASQ (American Society for Quality) propone una curiosa definición de calidad. Para ASQ [ASQ, 2014] "la calidad es una combinación de perspectivas cualitativas y cuantitativas para la que cada persona tiene su propia definición". Por ejemplo: "satisfacer los requisitos y expectativas que un servicio o producto debe cumplir" o "la persecución de soluciones óptimas que contribuyan a confirmar el éxito". En un contexto técnico, la calidad suele tener dos significados. El primero de ellos: "las características de un producto o servicio que le confieren su aptitud para satisfacer necesidades explícitas o implícitas". El segundo de ellos: "un producto o servicio sin deficiencias".

- Según Joseph Juran (consultor de gestión del siglo XX, principalmente recordado como un experto de la calidad y la gestión de la calidad), calidad significa "idoneidad para el uso".

- Para Philip Crosby (empresario estadounidense que contribuyó a la teoría gerencial y a las prácticas de la gestión de la calidad), calidad significa "conformidad con los requisitos".
- Según International Organization for Standardization (ISO) en el estándar ISO 9000:2005 [ISO, 2005a]: "Grado con el que un conjunto de características inherentes cumplen los requisitos". El estándar define los requisitos como necesidades o expectativas.
- Según la metodología Six Sigma, la calidad es el "número de defectos por millón".
- Otra definición interesante es la proporcionada por Peter Drucker (considerado uno de los padres de la disciplina del management): "La calidad de un producto o de un servicio no es lo que el proveedor pone en él. Es lo que el cliente obtiene y por lo que está dispuesto a pagar" [Drucker, 1985].
- En el contexto de los servicios de telecomunicación, ITU también ofrece varias definiciones de calidad en algunas de sus recomendaciones:
  - ITU-T E.800 [ITU, 2008a]: El conjunto de características de una entidad que le confieren su aptitud para satisfacer necesidades explícitas o implícitas.
  - ITU-T E.802 [ITU, 2007]: El conjunto de características de una entidad que le confieren su capacidad para satisfacer necesidades explícitas e implícitas. Estas características deben ser observables o medibles. Cuando se definen dichas características, éstas se convierten en parámetros y los parámetros se expresan mediante medidas.

Aunque la variedad de definiciones de calidad es grande, hay ciertos aspectos que comparten la mayoría de ellas. En primer lugar, la calidad es un aspecto inherente al objeto en cuestión, no es un añadido ni algo que se pueda añadir una vez creado el objeto. Además, varias definiciones coinciden en que la calidad de un objeto depende de las propiedades o las características del mismo. Por otro lado, las definiciones coinciden en destacar la calidad como la capacidad de que el objeto en cuestión realice adecuadamente las funciones para las que está diseñado. Sin embargo, la métrica para evaluar el desempeño del objeto varía entre definiciones (requisitos, expectativas, porcentaje de defectos, etc.).

Así pues, las definiciones presentadas ponen de manifiesto dos puntos de vista. Algunas definiciones asocian la calidad a una serie de parámetros observables y medibles, mientras que otras definiciones hablan de expectativas, de utilidad para el usuario, etc., las cuales son magnitudes más complejas de observar y medir. Estos dos puntos de vista han dado lugar a dos conceptos muy extendidos en el ámbito de la gestión de la calidad,

como son “calidad de servicio”, QoS y “calidad percibida” o “calidad de experiencia de usuario”, QoE, los cuales son tratados con más detalle en las siguientes secciones.

2.2. Calidad de servicio

2.2.1. Definiciones

En primer lugar, se presentan una serie de definiciones de calidad de servicio.

- Definiciones de ITU:
  - ITU-T E.800 [ITU, 2008a]: El conjunto de características de un servicio de telecomunicaciones que le confieren su aptitud para satisfacer las necesidades del usuario del servicio, ya sean explícitas o implícitas. Como se puede ver, la definición de calidad de servicio de ITU-T E.800 es simplemente la adaptación de su definición de calidad genérica a un servicio de telecomunicaciones.
  - ITU-T E.802 [ITU, 2007]: Esta recomendación ofrece dos definiciones complementarias. La primera de ellas es la misma definición que la ofrecida en la recomendación anterior. La segunda define la calidad de servicio como el efecto colectivo del rendimiento del servicio, que determina el grado de satisfacción de los usuarios del mismo.
  - ITU-T X.902 [ITU, 2009]: Un conjunto de cualidades relacionadas con el comportamiento colectivo de uno o más objetos. La QoS se puede especificar en un contrato y debe poder ser medida y reportada. La calidad de servicio está relacionada con características como la tasa de transferencia de información, latencia, probabilidad de que una comunicación se interrumpa, probabilidad de que el sistema falle, probabilidad de que el almacenamiento falle, etc.
- IETF RFC 2386 [Internet Engineering Task Force (IETF), 1998a]: El conjunto de requisitos de servicio que debe cumplir la red cuando transporta un flujo de datos.
- 3GPP TS 22.105 [3GPP, 2013]: El efecto colectivo de factores de rendimiento del servicio que determinan el nivel de satisfacción del usuario de un servicio.

En las recomendaciones ITU que se han comentado se definen varios puntos de vista de QoS: requisitos de QoS del usuario, QoS ofrecida por el proveedor del servicio, QoS conseguida/entregada por el proveedor del servicio y QoS percibida por el usuario. Como se puede ver, este último punto de vista está más próximo al concepto de QoE que de QoS.

2.2.2. Parámetros de rendimiento

Tradicionalmente, la gestión de la QoS está relacionada con la identificación de un conjunto de parámetros objetivos y medibles y la determinación de un conjunto de valores aceptables para los mismos. En [ITU, 2004b] se recopilan diversos parámetros de calidad de servicio que han sido identificados como claves en distintos estándares para la provisión de servicios de telecomunicaciones:

- Éxito en la conexión de la llamada
- Retardo en la conexión de la llamada
- Calidad conversacional y vocal
- Calidad en transmisiones por fax
- Métricas comparativas para rutas alternativas
- Calidad en transmisiones de vídeo
- Parámetros de error en la red de transporte
- Parámetros de rendimiento en redes IP

De especial relevancia en esta tesis son los parámetros de QoS utilizados en redes IP, por lo que se incluyen dos conjuntos de parámetros de rendimiento de redes IP, propuestos por el IETF y por ITU.

El grupo de trabajo IPPM (IP Performance Metrics) del IETF ha propuesto un conjunto de métricas de rendimiento en diversas RFC:

- Métricas de Conectividad (IP Connectivity Metrics), RFC 2678.
- Métrica de Retardo en Un Sentido (unidireccional) (One Way Delay Metric - OWD), RFC 2679.
- Métrica de Pérdida de Paquetes en Un Sentido (unidireccional) (One-Way Packet Loss Metric - OWPL), RFC 2680.
- Métrica de Retardo de Ida y Vuelta (bidireccional) (Round-Trip Delay Metric - RTD), RFC 2681.
- Métrica de Variación del Retardo de Paquetes (unidireccional) (IP Packet Delay Variation Metric - IPDV) (jitter), RFC 3393.
- Métrica de Capacidad de Transferencia (Bulk Transfer Capacity Metric) = Ancho de Banda Efectivo, RFC 3148.

- Métrica Muestral de Patrón de Pérdidas en Un Sentido (unidireccional) (One-Way Loss Pattern Sample Metric), RFC 3357.
- Métrica de Duplicación de Paquetes en Un Sentido (One-Way Packet Duplication Metric), RFC 5560.
- Métricas de Reordenamiento de Paquetes (Packet Reordering Metrics), RFC 4737.
- Métricas de Episodios de Pérdidas (Loss Episode Metrics), RFC 6534.
- Pérdidas de Ida y Vuelta (Round-Trip Packet Loss Metrics), RFC 6673.
- Métricas de Capacidad de la Red (Network Capacity), RFC 5136.
- Métricas de Rendimiento de la Red (Network Performance), RFC 3432.
- Métricas de Capacidad del Protocolo de Transporte (TCP Throughput), RFC 6349.

Por otra parte, ITU en su recomendación ITU-T Y.1540 [ITU, 2011d] define una serie de parámetros de rendimiento, similares a los propuestos por IETF IPPM:

- Disponibilidad: IP service availability
- Retardo: IP packet transfer delay (IPTD)
- Variación del retardo (jitter): IP packet delay variation (IPDV)
- Tasa de pérdidas de paquetes: IP packet loss ratio (IPLR)
- Tasa de paquetes erróneos: IP packet error ratio (IPER)
- Tasa de paquetes espurios: Spurious IP packet ratio (SIPR)
- Tasa de paquetes re-ordenados: IP packet reordered ratio (IPRR)
- Tasa de paquetes con pérdidas severas de bloque: IP packet severe loss block ratio (IPSLBR)
- Tasa de paquetes duplicados: IP packet duplicate ratio (IPDR)
- Tasa de paquetes replicados: Replicated IP packet ratio (RIPR)
- Parámetros a nivel de flujo
- Parámetros de capacidad

2.2.3. Clases de tráfico y clases de servicio

Además de la identificación de los parámetros más relevantes para la calidad de servicio, es necesario disponer de mecanismos que aseguren que dichos parámetros se encuentran dentro un rango determinado. La principal técnica que se ha utilizado para lograr este propósito se basa en la definición de "clases de servicio" que recogen las características del tipo de tráfico que se quiere cursar y de "contratos de servicio" que establecen los valores concretos de cada parámetro de QoS relevante que se deben asegurar.

Históricamente, uno de los primeros avances para el establecimiento de calidad de servicio en redes de datos fue el desarrollo de la arquitectura de protocolos ATM. En ATM se definieron Clases de Tráfico, en función de las características de los servicios soportados y Clases de Servicio, que trataban de dar respuesta a estos requisitos en la red de transporte. Estos conceptos han servido como base para el desarrollo de arquitecturas de QoS en otros sistemas y tecnologías.

Tanto el 3GPP como ITU han identificado diferentes requisitos de QoS para distintas clases de tráfico.

El 3GPP en su especificación TS 22.105 [3GPP, 2013] define las clases de tráfico que se muestran en la tabla 2.1, las cuales deben ser provistas a los usuarios finales extremo a extremo. Estas clases de tráfico están definidas en base a las necesidades de retardo y de tolerancia a errores de los distintos servicios.

Tabla 2.1: Clases de tráfico según 3GPP

	Conversacional (retardo <1s)	**Interactivo** (retardo 1s aprox.)	**Streaming** (retardo <10s)	**Segundo Plano** (retardo >10s)
Tolerante a errores	Conversaciones de voz y video	Mensajería de voz	Streaming de audio y video	Fax
No tolerante a errores	Telnet, juegos interactivos	E-Commerce, Navegación Web	FTP, imágenes fijas, paging	Notificaciones de E-mail

Para cada una de las clases de tráfico la especificación detalla los valores que deben cumplir un conjunto de métricas de rendimiento como la tasa de bit, el retardo extremo a extremo, jitter y tasa de pérdidas de paquete.

Por su parte, ITU en la recomendación ITU-T G.1010 [ITU, 2011a] define una serie de clases de tráfico similares a las comentadas anteriormente. Estas clases de tráfico se recogen en la tabla 2.2. De manera análoga a la especificación del 3GPP analizada anteriormente, ITU-T G.1010 define un rango de valores de tasa de bit, retardo extremo a extremo, jitter y tasa de pérdidas de paquete para cada una de las clases de tráfico consideradas.

Tabla 2.2: Clases de tráfico según ITU

	Interactivo (retardo $\ll$ 1s)	**Pronta respuesta** (retardo 2s aprox.)	**Puntual** (retardo 10s aprox.)	**No crítico** (retardo $\gg$ 10s)
Tolerante a errores	Voz en conversación y vídeo	Mensajería vocal/vídeo	Audio/vídeo en tiempo real	Fax
No tolerante a errores	Telnet, juegos interactivos	E-Commerce, Navegación Web	FTP, imágenes fijas	Tráfico de fondo

Una vez identificadas las distintas clases de tráfico, se establecen clases de servicio las cuales imponen una serie de restricciones al rendimiento de la red, con el objetivo de dar soporte a las distintas clases de tráfico.

En la tabla 2.3, se describen las clases de servicio o clases de QoS que ITU establece en su recomendación ITU-T Y.1541 [ITU, 2011e].

Tabla 2.3: Clases de servicio según ITU

Parámetro	**Clase 0**	**Clase 1**	**Clase 2**	**Clase 3**	**Clase 4**	**Clase 5**
IPTD	100 ms	400 ms	100 ms	400 ms	1s	U
IPDV	50 ms	50 ms	U	U	U	U
IPLR	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	U
IPER	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	U

Se debe destacar que para cada parámetro de rendimiento de red se establece un tipo de objetivo distinto:

- IPTD (IP packet transfer delay): Límite superior sobre el IPTD medio.
- IPDV (IP packet delay variation): Límite superior sobre el cuantil $1-10^{-3}$ menos el IPTD mínimo.
- IPLR (IP packet loss ratio): Límite superior.
- IPER (IP packet error ratio): Límite superior.

De manera adicional a las clases de la tabla 2.3, la recomendación ITU-T Y.1541 define dos clases de QoS provisionales con el objetivo de acomodar nuevas necesidades de los usuarios, aumentando los requisitos en cuanto a tasa de pérdidas y de errores de paquete.

Por otro lado, como se muestra en la tabla 2.4, [ITU, 2011e] proporciona ejemplos de aplicaciones para cada clase de QoS y recomendaciones en cuanto a las técnicas y mecanismos a utilizar para implementar dichas clases de QoS.

Tabla 2.4: Clases de servicio según ITU: aplicaciones e implementación

Clase	Aplicaciones	Mecanismos de nodo	Técnicas de red
0	Tiempo real, sensible al jitter, alta interactividad (VoIP, VTC)	Cola independiente con servicio preferentes, acondicionamiento de tráfico	Encaminamiento y distancias restringidos
1	En tiempo real, sensible al jitter, interactividad normal (VoIP, VTC)	Cola independiente con servicio preferentes, acondicionamiento de tráfic	Encaminamiento y distancias menos restringidos
2	Datos de transacciones, alta interactividad (señalización)	Cola independiente, sin prioridades	Encaminamiento y distancias restringidos
3	Datos de transacciones, interactividad normal	Cola independiente, sin prioridades	Encaminamiento y distancias menos restringidos
4	Sólo de baja pérdida (transacciones cortas, datos en bloque, video de flujo continuo)	Cola larga, sin prioridades	Cualquier ruta
5	Aplicaciones tradicionales de las redes IP	Cola independiente, prioridad mínima	Cualquier ruta

2.2.4. Mecanismos de implementación de QoS en redes IP

Además de definir clases de servicio o clases de QoS, es necesario contar con tecnologías capaces de imponer las restricciones que dictan dichas clases. Aunque no son directamente aplicables a esta tesis, en esta sección se revisan brevemente algunos de los mecanismos de implementación de calidad de servicio en redes IP más destacados.

La primer técnica de cierta relevancia para proveer de QoS a las redes IP fue el modelo de servicios integrados (IntServ, Integrated Services). IntServ [IETF, 1994] se basa en la reserva de recursos de la red, dividiendo el tráfico en diferentes tipos de flujos. La implementación de IntServ requiere mantener un estado (soft state) por cada tipo de flujo en cada nodo de la red, además de un protocolo de señalización que permita gestionar la reserva de la red. Aunque la especificación no establece ningún protocolo concreto, el protocolo de reserva de recursos más usado es RSVP (Resource Reservation Protocol) [IETF, 1997]. Además del tratamiento best effort, IntServ ofrece los denominados "servicio garantizado" (garantiza los niveles solicitados para todos los parámetros de rendimiento) y "servicio de carga controlada" (no garantiza todos los parámetros de rendimiento pero ofrece baja tasa de pérdidas de paquetes).

El modelo de servicios diferenciados (DiffServ, Differentiated Services) se basa en marcar cada paquete que se envía por la red [IETF, 1998b]. En base a estas marcas, los nodos de la red deciden el trato que aplican a cada paquete. Esta solución es más

escalable que IntServ, ya que elimina la necesidad de mantener información en los nodos de la red por cada flujo cursado. Además del tratamiento best effort, DiffServ ofrece "assured forwarding" (similar al servicio de carga controlada de IntServ) y "expedited forwarding" (similar al servicio garantizado de IntServ).

Por último, aunque no es como tal una arquitectura de provisión de QoS, MPLS (Multiprotocol Label Switching) [IETF, 2001] permite forzar rutas para los paquetes que porten cierta etiqueta, por lo que se puede combinar con mecanismos de ingeniería de tráfico para implementar mecanismos de QoS.

2.3. Calidad percibida

En varias de las definiciones anteriores, tanto del concepto general de calidad, como del concepto de calidad de servicio, se deja entrever una perspectiva subjetiva, cercana al usuario final y relacionada con sus expectativas y el nivel de satisfacción que obtiene al consumir el servicio.

Por ejemplo, ITU-T E.800 [ITU, 2008a] introduce el concepto de QoSE (QoS experienced/perceived by customer/user) como el nivel de calidad que los clientes o usuarios creen que han experimentado. ITU-T E.802 [ITU, 2007] define la QoS como el efecto colectivo del rendimiento del servicio, que determina el grado de satisfacción de los usuarios del mismo. El 3GPP por su parte, define la calidad de servicio en su especificación TS 22.105 [3GPP, 2013] como el efecto colectivo de factores de rendimiento del servicio que determinan el nivel de satisfacción del usuario de un servicio.

Estas definiciones ponen de manifiesto dos planos de calidad íntimamente relacionados: el plano de la calidad de servicio, relacionado con el rendimiento de la red y el plano de la calidad percibida o calidad de experiencia QoE, una dimensión del concepto de calidad más amplia que incluye aspectos subjetivos relacionados con la percepción de los usuarios finales.

2.3.1. Definiciones

- Definiciones de ITU:
  - ITU-T P.10 [ITU, 2008e]: aceptabilidad general de una aplicación o servicio, percibida subjetivamente por los usuarios finales. La QoE incluye los efectos del sistema extremo a extremo (cliente, terminal, red, infraestructura, etc.). La aceptabilidad general puede estar influenciada por las expectativas y el contexto del usuario.
  - ITU-T E.800 [ITU, 2008a] e ITU-T G.1000 [ITU, 2001]: define un concepto de calidad de servicio percibida (QoSE) como el nivel de calidad que

los clientes o usuarios creen que han experimentado. Además introduce los siguientes conceptos:

- El nivel de calidad de servicio percibida se puede expresar como una escala de opinión.
- La QoSE tiene un componente cuantitativo y otro cualitativo. El componente cuantitativo puede estar influido por el efecto extremo a extremo del sistema. El factor cualitativo puede estar influenciado por las expectativas del usuario, condiciones ambientales, factores psicológicos, contexto de la aplicación, etc.

- ITU-T G.1010 [ITU, 2011a]: aunque no introduce específicamente el término de QoE, hace referencia a la importancia de especificar los requisitos de los servicios y aplicaciones desde el punto de vista del usuario. En concreto, según esta recomendación el rendimiento debe ser expresado mediante parámetros que:
  - Tengan en cuenta todos los aspectos del servicio desde el punto de vista del usuario.
  - Se centren en efectos perceptibles por el usuario y no tanto en las causas que los provocan.
  - Independientes de la tecnología y de la arquitectura de red.
  - Puedan ser objetiva o subjetivamente medidos.
  - Puedan ser fácilmente relacionados con parámetros de rendimiento de red.
  - Puedan ser asegurados al cliente por parte de los proveedores de servicio.

- En [Patrick Le Callet and Perkis, 2013] se propone la siguiente definición: calidad de experiencia o QoE es el grado de placer o disgusto del usuario de una aplicación o servicio. Es el resultado de la realización de sus expectativas con respecto a la utilidad y/o disfrute de la aplicación o servicio en función de la personalidad del usuario y de su estado actual.

Aparte de estas definiciones, existen numerosas otras que describen a grandes rasgos los mismos aspectos que en las anteriores. Quizás, de especial interés puede ser lo descrito en ETSI EG 202 765-1 [ETSI, 2010], donde se destacan los dos grandes problemas a la hora de evaluar de forma global la QoS de un servicio. El primer problema es la diferencia (gap) que existe entre los aspectos técnicos y los aspectos perceptivos y el segundo que tanto la QoS como la satisfacción global son difíciles de modelar, ya que dependen de forma importante de las expectativas y de otros aspectos contextuales y subjetivos.

En definitiva, comparando las definiciones anteriores se puede comprobar que casi todas tienen muchos puntos en común, por lo que se pueden realizar las siguientes afirmaciones:

1. La Calidad de Experiencia (QoE) depende de la percepción subjetiva de los usuarios.

2. La QoE tiene dos componentes:

   a) Objetivo, cuantitativo o tangible, que depende de la calidad de funcionamiento del sistema extremo-a-extremo (calidad técnica).

   b) Subjetivo, cualitativo o intangible, en el que influyen las expectativas del usuario, las condiciones ambientales, factores psicológicos y contextuales, etc. (calidad subjetiva).

3. La satisfacción de los usuarios depende de la diferencia (gap) entre los requisitos y expectativas de los usuarios (calidad requerida) y la calidad percibida o experimentada por dichos usuarios en la utilización del servicio.

2.3.2. Modelos generales de calidad percibida

El concepto de QoE no es un concepto que se aplique únicamente en la industria de las telecomunicaciones sino que se aplica en multitud de técnicas de gestión y mejora de procesos productivos en diversos ámbitos. En esta sección se describen algunos modelos generales de representación de la calidad percibida.

2.3.2.1. Modelo de Oliver

Este modelo propone que la evaluación de la calidad percibida es el resultado de las discrepancias entre las expectativas y las percepciones de los usuarios sobre el funcionamiento de un servicio. Como se muestra en la figura 2.1, dicho modelo se apoya en el "Paradigma de la Expectativa-Disconfirmación" [Oliver, 2009], de modo que la satisfacción es el resultado del cumplimiento de las expectativas y la insatisfacción que se produce cuando éstas no se cumplen.

2.3.2.2. Modelo de Grönroos

El modelo propuesto por Grönroos [Grönroos, 1984] identifica varios factores críticos que afectan a la evaluación de la calidad en la prestación de un servicio:

1. La calidad técnica, determinada por las características inherentes al servicio.

2. La calidad funcional o relacional, determinada por la forma en que se presenta el servicio.

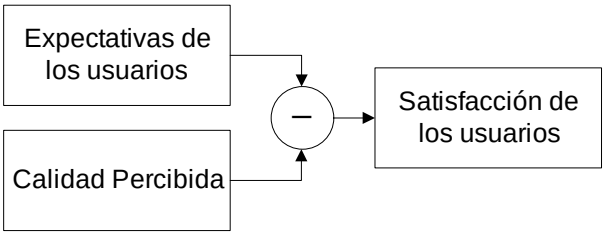


Figura 2.1: Modelo de QoE de Oliver

3. La impresión o percepción por parte del cliente acerca del proveedor, determinada por la imagen de la empresa, las experiencias previas y otros factores.

En este modelo se presenta como factor más importante la calidad funcional, siempre que la calidad técnica supere un cierto umbral mínimo. Además, este modelo considera que los usuarios realizan la evaluación de la calidad comparando el servicio esperado con el servicio recibido.

Asimismo, en este modelo también se definen una serie de criterios que afectan a la calidad del servicio:

- Pericia profesional
- Actitud y conducta
- Cercanía y flexibilidad
- Fiabilidad
- Soporte a situaciones de error
- Imagen de marca

2.3.2.3. Modelo SERVQUAL

El modelo SERVQUAL [Parasuraman et al., 1988] [Parasuraman et al., 1991] describe la calidad del servicio como un concepto abstracto, debido a que el servicio es algo intangible, heterogéneo e inseparable.

Este modelo hace una clara distinción entre Calidad Esperada y la Calidad Percibida, a partir de cuatro factores que determinan la ausencia de calidad:

- La ignorancia de las expectativas del cliente por parte del proveedor.
- La falta de normas.
- La discordancia entre las normas.

- El incumplimiento de las promesas realizadas por el proveedor.

Este modelo define la Calidad de Servicio Percibida como la diferencia entre las expectativas de los usuarios y la percepción de los mismos acerca del servicio recibido. Se identifican algunos factores que contribuyen a esa diferencia debido a posibles "desajustes" o carencias en la cadena de provisión de un servicio:

- Diferencia entre las expectativas del cliente y la percepción de las mismas por el proveedor del servicio.
- Diferencia entre la percepción de las expectativas por el proveedor y su traducción a requisitos o especificaciones de calidad.
- Diferencia entre la calidad de servicio especificada y la realmente implementada o entregada.
- Diferencia entre el servicio prestado y el ofertado al cliente.
- Diferencia entre las expectativas del usuario y las características del servicio percibido por el usuario.

El modelo establece que la diferencia final, es decir, la diferencia entre expectativas del servicio y las características del servicio percibido, es función de las anteriores diferencias.

El modelo intenta medir estas expectativas y la percepción de los usuarios mediante una encuesta de 22 preguntas clasificadas en categorías o "dimensiones" que se consideran comunes a todos los servicios:

1. Elementos tangibles: Instalaciones, equipamiento y apariencia del personal.
2. Fiabilidad: Eficacia en la prestación del servicio.
3. Capacidad de respuesta: Rapidez en la respuesta a las consultas y/o quejas de los usuarios.
4. Garantía: Competencia, cortesía, credibilidad y seguridad.
5. Empatía: Incluye la capacidad del cliente de utilizar el servicio cuando lo desee (acceso), la habilidad para informarle en su propio lenguaje (comunicación) y el conocimiento de sus necesidades y expectativas.

El nivel de importancia de cada una de las cinco dimensiones depende tanto del tipo de servicio ofrecido como del valor que cada una implica para el cliente, lo que se verá reflejado directamente en los resultados de las encuestas.

2.3.2.4. Modelo SERVPERF

El modelo SERVPERF [Cronin and Taylor, 1992], [Cronin and Taylor, 1994] es una variante del modelo SERVQUAL que se basa en la idea de que el concepto de Calidad de Servicio basado en las diferencias entre expectativas y percepciones de los usuarios es inadecuado, ya que postula que no existe justificación teórica suficiente para que estas magnitudes sirvan para la medición de la calidad de servicio.

Por esta razón, el modelo SERVPERF propone basarse únicamente en la percepción de los usuarios de los servicios. Asimismo, pretende analizar las relaciones entre calidad de servicio, satisfacción del cliente e intenciones de compra.

Las características del modelo SERVPERF son las siguientes:

- Sólo tiene en cuenta las percepciones del usuario con respecto al servicio (ignorando sus expectativas acerca del mismo).
- Consta de 22 elementos clasificados en cinco dimensiones, de forma similar a SERVQUAL.
- Utiliza un único formulario (escala) para las preguntas, ya que no pretende medir las expectativas.

La escala y las dimensiones son las mismas que en el modelo SERVQUAL. No obstante, el enfoque de evaluación varía que ya que sólo se tiene en cuenta las percepciones del usuario. La principal desventaja del modelo es que debido a no tener en cuenta los requisitos y las expectativas de los usuarios, no se pueden establecer qué características o aspectos son necesarios mejorar.

2.3.2.5. Modelo de Hardy

El modelo de Hardy [Hardy, 2001] identifica tres componentes diferentes de calidad: la QoS intrínseca, la QoS percibida por el usuario y la valoración global de QoS:

1. La QoS intrínseca se relaciona con el concepto de calidad de funcionamiento de red, es decir, parámetros y métricas de rendimiento de red (retardo, variación del retardo, pérdidas, caudal, etc.).
2. La QoS percibida hace referencia a la calidad tal y como la experimenta el usuario.
3. La valoración global de la QoS se refiere al grado de satisfacción del usuario con el servicio y su intención para volver a contratarlo con el mismo proveedor.

En la figura 2.2 se muestra un diagrama con las dimensiones de la QoS según el modelo de Hardy.

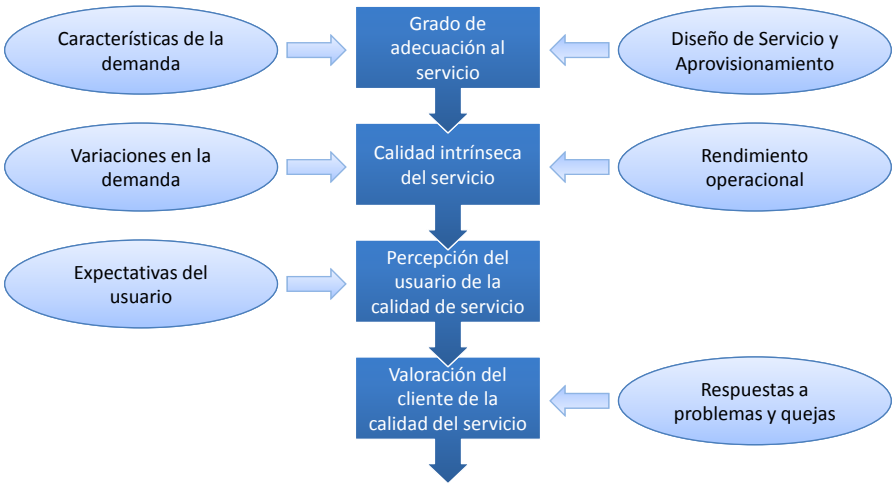


Figura 2.2: Modelo de calidad de Hardy

2.3.3. Medida de QoE en servicios de telecomunicaciones

En las secciones anteriores se ha puesto de manifiesto la complejidad que supone definir unos niveles de calidad técnica o calidad de servicio que garanticen un cierto nivel de calidad percibida por los usuarios, ya que existen una serie de factores significativos para la evaluación del servicio por parte de los usuarios que van más allá de parámetros puramente técnicos. Esto conlleva que aunque se satisfagan los niveles requeridos en cuanto a rendimiento técnico, no siempre se consiga un nivel de calidad percibida adecuado. Así pues, en la literatura se han definido métodos que permiten evaluar la calidad percibida, teniendo en cuenta su componente subjetiva, además de intentar relacionar la calidad técnica o de servicio, QoS, con la calidad subjetiva o calidad percibida, QoE.

2.3.3.1. Mean Opinion Score (MOS)

El método de evaluación subjetiva de calidad MOS se basa en el análisis de las opiniones de los usuarios respecto a un servicio. Cada cliente evalúa el servicio recibido mediante una única calificación, por lo que en dicha evaluación están incluidos diferentes aspectos, tanto objetivos como subjetivos. La escala de calificación comprende valores entre 1 (calidad inaceptable) y 5 (calidad excelente). El valor de MOS se obtiene realizando la media aritmética de las calificaciones de todos los usuarios.

El método de evaluación MOS, definido en [ITU, 1997b] fue concebido inicialmente para evaluar la opinión de los clientes de telefonía en cuanto a la calidad auditiva. Refleja directamente la calidad percibida por los usuarios, por lo que se ha utilizado

ampliamente en entornos controlados tanto para la evaluación de la calidad de las líneas de transmisión como de los algoritmos de codificación de la señal de voz (codecs) en líneas digitales.

Este método de medida de calidad ha demostrado ser muy fiable, por lo que ha sido adoptado en diversos ámbitos como herramienta de medida de calidad percibida. Por ejemplo, en [ITU, 2008f], [ITU, 1998c] y [ITU, 1998d] se utiliza este método en aplicaciones multimedia.

Sin embargo, a pesar de su fiabilidad, este método presenta algunos inconvenientes.

- No permite identificar qué aspectos del servicio han llevado a los usuarios a evaluar el servicio negativamente.
- Es costoso en recursos y en tiempo.
- No se puede aplicar directamente a la medición de la calidad en tiempo real.

2.3.3.2. Evaluación continua

Existen ciertos entornos en los que la evaluación de la calidad en momentos puntuales no es suficiente, debido principalmente a la variabilidad del entorno. Por ejemplo, en la evaluación de la calidad de un servicio que se presta a través de Internet no basta con realizar una medida de calidad en un instante de tiempo concreto, ya que la propia variabilidad de la red puede hacer que en otro instante las condiciones en las que se presta el servicio cambien drásticamente, modificándose por tanto el nivel de calidad del servicio.

Por tanto, para entornos altamente variables se necesitan técnicas o instrumentos que permitan la evaluación de la calidad de manera continuada. En [Bouch and Sasse, 1999] se llevan a cabo una serie de evaluaciones de calidad en sesiones de audio interactivo utilizando una herramienta denominada QUASS (Quality Assessment Slider), la cual implementa un control deslizable con el que los usuarios pueden dar una valoración continua de la calidad que perciben. ITU también define metodologías continuas de evaluación de calidad en sus recomendaciones ITU-T P.910 [ITU, 2008f] e ITU-T P.911 [ITU, 1998c].

2.3.3.3. Métodos de estimación

Estos métodos se basan en la estimación de la calidad percibida a partir de medidas de rendimiento. Son métodos objetivos que tratan de modelar de manera cuantitativa de las relaciones entre la calidad percibida por los usuarios finales y las medidas objetivas de determinados parámetros de calidad. Los modelos se determinan a partir de experimentos realizados con usuarios, en los que se recogen sus percepciones sobre la calidad experimentada en diversas condiciones.

La metodología para el desarrollo de estos modelos tiene generalmente tres fases:

1. Estudios empíricos para obtener valoraciones de los usuarios finales respecto de uno o más servicios en concreto. Los resultados dependerán, en general, del tipo o perfil de usuarios.

2. Definición de modelos de estimación de la calidad percibida por los usuarios finales en función de parámetros objetivos de calidad elegidos.

3. Definición de métodos de medida de los parámetros de interés.

Estos métodos permiten la evaluación continua de la QoE de un servicio a partir de medidas objetivas del rendimiento, QoS. Su principal desventaja es que dependen fuertemente de los servicios asociados, por lo que los parámetros críticos de calidad varían de un servicio a otro.

Este tipo de métodos son de especial relevancia en esta tesis, ya que los modelos de calidad percibida que se van a desarrollar a lo largo de la misma se pueden considerar métodos de estimación de calidad, tal como se definen en esta sección. Los modelos desarrollados ofrecen una estimación de la calidad percibida por los usuarios a partir del análisis de un conjunto de parámetros de rendimiento y otros parámetros objetivos del servicio. Además, para el desarrollo de algunos de estos modelos se seguirá la metodología anterior, realizando una serie de experimentos de evaluación subjetiva de calidad con los que se obtendrán datos que permitirán entender mejor el proceso humano de generación de valoraciones de calidad y el planteamiento de expresiones matemáticas que modelen dichas valoraciones.

En la literatura se han propuesto un gran número de modelos de estimación de calidad, orientados a diferentes servicios y utilizando una gran variedad de parámetros y de técnicas en su desarrollo. En esta tesis se han revisado diversos trabajos relacionados con el ámbito de interés de la misma y se ha decidido describir este análisis en el capítulo correspondiente. Así pues, se remite al lector a las siguientes secciones si desea profundizar más en los trabajos relacionados con los siguientes aspectos:

- Modelos de estimación de calidad audiovisual para flujos sincronizados: sección 3.4.1.1.

- Efecto de la sincronización audio-vídeo en la calidad percibida: sección 3.4.2.

- Modelos de estimación de calidad asociada al tiempo de cambio de canal: sección 3.5.1.1.

- Modelos de estimación de calidad asociada a la función de acceso aleatorio en vídeo: sección 3.5.2.1.

- Modelos de estimación de calidad de vídeo: sección 4.2.
- Modelos de degradación de calidad debida a la transmisión (efecto de la red) para vídeo OTT: sección 5.2.

2.4. MPEG-DASH

2.4.1. Introducción

Como se ha comentado en el capítulo de introducción, en los últimos años se ha experimentado un notable aumento en el consumo de servicios de vídeo sobre Internet. Según datos de "The Diffusion Group (TDG) Research", en enero de 2014 el 63 % de los hogares norteamericanos tenían al menos una televisión conectada a Internet (ya sea smart TV o una televisión conectada a través de otro dispositivo). Una encuesta realizada por GfK en septiembre de 2013 descubrió que el 51 % de la población de Estados Unidos, con edades comprendidas entre los 13 y los 54 años, veían programas de televisión o películas mediante streaming de vídeo, al menos una vez a la semana. Es especialmente interesante la evolución de estos datos, ya que la misma encuesta reveló que dicho porcentaje era del 37 % y del 48 % en 2010 y 2012 respectivamente.

Sin embargo, existen algunas razones que están impidiendo que los servicios de streaming de vídeo OTT alcancen todo su mercado potencial. Una de estas razones es que la mayoría de las plataformas comerciales de vídeo OTT que existen actualmente son sistemas cerrados, con sus propios protocolos, formatos de descripción y representación de contenidos, etc., es decir, no existe interoperabilidad entre servidores y dispositivos de distintos fabricantes u operadores.

Desde un punto de vista más técnico, se puede decir que desde sus inicios, en torno al año 1990, la distribución de vídeo a través de Internet se encontró con dos problemas principales:

1. La realización de la entrega de contenidos a tiempo.
2. El coste asociado al envío de grandes cantidades de datos.

Para intentar resolver el primer problema, el IETF diseñó el protocolo Real-time Transport Protocol (RTP), el cual define formatos de paquete y mecanismos de control de sesión para realizar streaming multimedia en redes IP. Aunque RTP funciona bien en redes IP gestionadas, presenta algunos problemas cuando se usa a través de Internet. En primer lugar, muchas CDN no soportan RTP, por lo que "acercar" físicamente el contenido a los usuarios se presenta como un reto. Por otro lado, los paquetes RTP no suelen ser aceptados por los firewalls. Además, el diseño de RTP hace que el servidor tenga que gestionar información de sesión para cada usuario de manera independiente.

Todos estos problemas hace que desplegar una solución basada en RTP de manera masiva sea todo un reto tecnológico.

Con el paso de los años, el segundo problema se ha ido reduciendo, ya que el incremento de los anchos de banda ha reducido bastante el coste que supone el envío de información a través de Internet. Este hecho, junto con el enorme crecimiento de la World Wide Web, hace que la distribución de contenido multimedia pueda realizarse de manera eficaz enviando fragmentos de vídeo (segmentos) a través del protocolo Hypertext Transfer Protocol (HTTP). El streaming basado en HTTP tiene las siguientes ventajas:

- La infraestructura de Internet ha evolucionado para adaptarse de manera eficaz al tráfico HTTP. El ejemplo más importante de esta adaptación son las CDN, que proporcionan réplicas del contenido en localizaciones cercanas al usuario para reducir el tráfico en las redes troncales.
- HTTP atraviesa la mayor parte de firewalls ya que suelen estar configurados para soportar conexiones HTTP salientes.
- La tecnología de servidores HTTP es muy barata.
- Mediante streaming HTTP son los clientes los que mantienen la información de sesión, por lo que la escalabilidad de este tipo de servicios es muy alta.

Estas ventajas han propiciado la aparición de diferentes plataformas de streaming como HTTP Live Streaming (Apple), Smooth Streaming (Microsoft), HTTP Dynamic Streaming (Adobe), cada una con diferentes formatos de segmentos y diferentes ficheros de manifiesto, por lo que la interoperabilidad entre ellas no es inmediata.

Volviendo al punto anterior, esta falta de interoperabilidad entre plataformas es una de las principales motivaciones que han llevado al desarrollo del estándar que se describen a continuación, MPEG-DASH.

2.4.2. Streaming adaptativo

Antes de entrar en detalles concretos del estándar MPEG-DASH es importante tener una noción del concepto de streaming adaptativo sobre HTTP.

Como se comentó en la introducción, en los sistemas de streaming sobre HTTP es el cliente el que mantiene el control de la sesión. En este contexto, mantener la sesión significa, entre otras cosas, solicitar al servidor los fragmentos de vídeo necesarios para la reproducción del contenido. Para ello, es necesario algún mecanismo que permita a los clientes conocer qué fragmentos están disponibles en el servidor y cómo solicitarlos. En general, los sistemas de streaming de vídeo HTTP resuelven esta cuestión poniendo

en el servidor un fichero a disposición de los clientes con la información necesaria para que éste lleve a cabo las peticiones necesarias para obtener los fragmentos de vídeo. A estos ficheros se les suele conocer como ficheros manifest, aunque en MPEG-DASH se les denomina ficheros Media Presentation Description (MPD).

En los sistemas de streaming adaptativo, se incluye información en el fichero manifest sobre un catálogo de versiones disponibles en el servidor para un mismo contenido. Por ejemplo, diferentes representaciones del flujo de vídeo codificado a distintas tasas de bit, audio en diferentes idiomas, etc.. Una vez que el cliente conoce las distintas versiones del contenido nada le impide, en cada petición, conmutar entre ellas. Dicha conmutación puede realizarse como respuesta a una acción del usuario (por ejemplo, cambiar el idioma del audio) o bien, caso de uso típico de estos sistemas, como respuesta a un cambio en las condiciones de la red (por ejemplo, si la tasa de bit disponible en la red se reduce, la aplicación cliente puede decidir conmutar a un nivel de calidad inferior, solicitando segmentos de vídeo codificado a menor tasa de bit que la actual).

2.4.3. Arquitectura de referencia y alcance del estándar

En la figura 2.3 se presenta la arquitectura general de un servicio basado en MPEG-DASH. Como se puede ver, los bloques que componen la figura implementan un servicio de streaming adaptativo, de acuerdo a la descripción del apartado anterior. El servidor HTTP almacena los segmentos de los distintos flujos de medios y el fichero de descripción MPD, mientras que el cliente consta de un motor encargado de realizar las peticiones de segmentos y de una serie de módulos que permiten decodificar y renderizar el contenido de cada segmento.

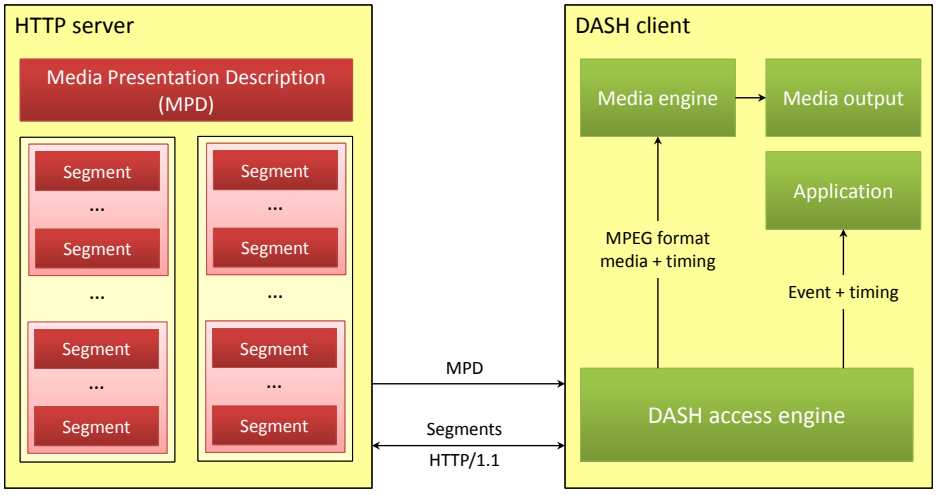


Figura 2.3: Arquitectura genérica de MPEG-DASH

La última versión del estándar MPEG-DASH se denomina ISO/IEC 23009:2014 y está formado por cuatro partes, siendo la más relevante la primera de ellas:

1. Media presentation description and segment formats
2. Conformance and reference software
3. Implementation guidelines (Technical Report)
4. Segment encryption and authentication

En la primera parte del estándar (Media presentation description and segment formats) [ISO, 2014b] se establece el formato que deben seguir tanto el fichero MPD como los segmentos que en él se definen. El protocolo que el estándar propone para la transmisión de segmentos es HTTP/1.1. Es importante destacar que el estándar solo define el formato del fichero MPD y el formato de los segmentos. La transmisión del MPD y el comportamiento del cliente en cuanto a reproducción y mecanismos de adaptación quedan fuera del estándar.

2.4.4. Estructura del fichero MPD

Como se ha comentando anteriormente, el estándar MPEG-DASH no se centra en procedimientos propios de cliente o servidor, sino que pone el foco en el formato de los segmentos y del fichero MPD que los describe. Este fichero está formateado en XML y describe el contenido que el servidor pone a disposición de los clientes, además de informales de cómo deben solicitar dicho contenido.

En la figura 2.4 se muestra de manera esquemática la estructura típica de un fichero MPD.

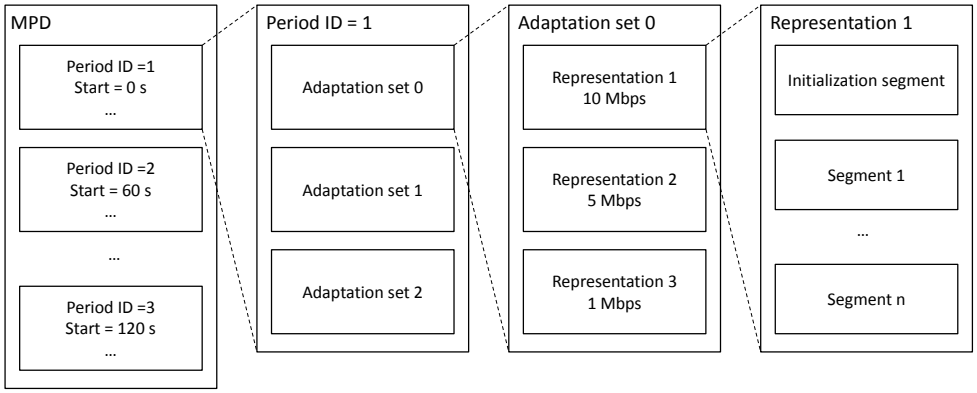


Figura 2.4: Estructura del MPD de MPEG-DASH

El contenido multimedia descrito por este MPD se compone de uno o varios **periodos** contiguos en el tiempo. Un periodo representa un intervalo de tiempo del contenido multimedia en el que el conjunto de versiones codificadas permanece constante. Dentro de un periodo, el contenido se organiza en **conjuntos de adaptación** o adaptation sets, los cuales representan un conjunto intercambiable de versiones codificadas de un componente (audio, vídeo, etc.) del contenido multimedia. Por ejemplo, un conjunto de adaptación puede contener diferentes versiones de la componente de vídeo codificada a varias tasas de bit, mientras que otro conjunto de adaptación puede estar formado por diferentes versiones de la componente de audio en diversas calidades.

Un conjunto de adaptación contiene un conjunto de **representaciones** que describen una versión codificada de uno (o varios, si están multiplexados) componentes del contenido multimedia. Por tanto, los clientes puedes conmutar dinámicamente entre representaciones de un mismo conjunto de adaptación para adaptarse a las condiciones de la red o a otros factores.

En cada representación, el contenido se divide en el tiempo en forma de **segmentos**, los cuales están identificados por una URL. Los segmentos se puede dividir a su vez en **subsegmentos**, cada uno de los cuales contiene un número entero de unidades de acceso.

2.4.5. Formato de los segmentos

El estándar MPEG-DASH es independiente del tipo de codificación de los contenidos y define formatos para contenedores de segmentos ISO Base Media File Format [ISO, 2005b] y MPEG-2 Transport Stream [ISO, 2013a]. Además, ofrece recomendaciones para realizar extensiones a otros formatos.

2.5. Codificación de vídeo

2.5.1. Introducción

La mayoría de los codecs de vídeo estándar siguen el mismo principio de funcionamiento, basado en tres principios de reducción de redundancia [Ghanbari, 2003]:

1. Reducción de redundancia espacial: para reducir la redundancia espacial se suelen utilizar técnicas como la codificación en dominios transformados, codificación predictiva, etc.

2. Reducción de redundancia temporal: codificación de diferencias entre entre imágenes sucesivas, estimación y compensación del movimiento, etc.

3. Codificación basada en entropía: utilización de códigos de longitud variables para reducir la redundancia entre símbolos.

De estas técnicas, la que ha conseguido mejores resultados en cuanto a tasas de compresión es la predicción y compensación de movimiento, a cambio, no obstante, de un importante incremento en la complejidad computacional.

En el resto de esta sección se ofrece una revisión histórica de la evolución de los principales codecs de vídeo y se presenta una introducción al proceso de codificación utilizado en H.264.

2.5.2. Evolución de los estándares de codificación de vídeo

En esta sección se lleva a cabo una revisión de los principales estándares de codificación de vídeo. La estandarización de sistemas de codificación de vídeo ha estado dominada tradicionalmente por dos organismos:

- ITU-T Video Coding Experts Group (VCEG)
- ISO/IEC Moving Picture Experts Group (MPEG)

H.120 [ITU, 1993a] fue el primer estándar de compresión de vídeo digital, desarrollado en 1984 por COST 211 y publicado por CCIT (actual ITU-T). Este codec estaba basado en DPCM (Differential Pulse Code Modulation), cuantificación escalar y reutilización de zonas comunes entre tramas. Se desarrolló una segunda versión en 1988 que incluyó compensación de movimiento y predicción de fondo. Aunque este codec no fue lo suficientemente bueno para ser utilizado de manera práctica, contribuyó a plantear la idea de que para conseguir unos ratios de compresión aceptables era preciso bajar la barrera del bit por pixel, dando lugar a la codificación por bloques que se utilizó en codificadores posteriores, como H.261.

H.261 [ITU, 1993b] se considera la base de los estándares modernos de compresión de vídeo. Aunque se desarrolló a finales de 1990, su estructura es similar a la de codecs actuales: estimación y compensación de movimiento en macrobloques de 16x16 píxeles, DCT de 8x8 píxeles, escaneo de coeficientes DCT en zigzag, cuantificación escalar de coeficientes DCT y codificación de longitud variable. Fue diseñado para ser utilizado en RDSI, por lo que soporta tasas de codificación múltiplo de 64 kbit/s.

El codificador de vídeo MPEG-1 se definió en ISO/IEC 11172 parte 2 en 1993 [ISO, 1993]. Fue el primer estándar de codificación de vídeo desarrollado por ISO. Se diseñó principalmente para aplicaciones de almacenamiento de vídeo y utiliza una estructura similar a H.261, pero introduce nuevos conceptos como: predicción bidireccional (tramas B), codificación mediante slices, matrices de pesos de cuantificación, etc.

Entre el año 1994 y 1995 una iniciativa conjunta de ISO y de ITU-T dio lugar al estándar ISO/IEC 13818-2 (MPEG-2) [ISO, 2013b] o H.262 [ITU, 2012d]. Este estándar se utiliza ampliamente en DVD y en televisión distribución de TV (DVB). Sus

principales novedades son el soporte para imágenes entrelazadas, incremento en la precisión de la cuantificación, diversas formas de escalabilidad y utilización de vectores de movimiento en tramas I para corrección de errores. Fue diseñado para aplicaciones con altas tasa de bit (2-20 Mbps) y no es adecuado para aplicaciones con menos de 1 Mbps.

El siguiente paso en la carrera de los codificadores de vídeo vino de la mano de H.263 en 1995 [ITU, 2005]. Este nuevo estándar fue superior a todos sus predecesores a cualquier tasa de bit (excepto en vídeo entrelazado) y especialmente a bajas tasas de bit (superior en un factor de 2). Algunas de las principales novedades que incorpora son: codificación de longitud variable de los coeficientes de la DCT 3D, mejora de la predicción de vectores de movimiento, tramas PB (dos tramas P y B se codifican como una única entidad), etc. En 1998 y en 2000 se presentaron dos nuevas versiones de H.263, conocidas como H.263+ y H.263++ las cuales mejoraron aspectos como la tolerancia a errores y la escalabilidad, con el objetivo de adaptarse a las nuevas aplicaciones móviles y de Internet.

La primera versión de MPEG-4 parte 2 (ISO/IEC 14496-2) [ISO, 2004] apareció a principios de 1999, incluyendo las mismas características que H.263 e incluyendo funciones típicas de VCR (trick modes). MPEG-4 parte 2 es más eficiente que H.263, especialmente a bajas tasas de bit. Incluye novedades con respecto a H.263, entre las que destacan: mejora de la tolerancia a errores, codificación de varios objetos en la misma trama, codificación de formas, codificación wavelet de imágenes fijas, etc. Existen diferentes perfiles, aunque no todos han sido implementados. En 2000 y en 2001 se desarrollaron dos nuevas versiones de este codificador que incluyen nuevas funcionalidades de compensación de movimiento. A pesar de la cantidad de mejoras de MPEG-4, éste no fue especialmente adoptado por los fabricantes, seguramente por el cambio de paradigma que algunas de sus funcionalidades suponen, pasando de una codificación basada en macrobloques a una codificación basada en objetos.

Aunque comenzó a desarrollarse a mediados de 1999, no fue hasta 2003 cuando se completó la estandarización de ITU-T H.264/AVC (Advanced Video Coding) [ITU, 2014c] o MPEG-4 parte 10 (ISO/IEC 14496-10) [ISO, 2014a]. H.264 soporta un amplio rango de resoluciones y tasas de bit, por lo que es adecuado para múltiples aplicaciones como distribución de vídeo, almacenamiento, transmisión por redes de paquetes, etc. Es más complejo que sus predecesores pero consigue mayores tasas de compresión gracias a funcionalidades (algunas de ellas extraídas de H.263++) como codificación con predicción intra-trama, compensación de movimiento multi-trama y de tamaño de bloque variable, precisión en la estimación desde un cuarto a un octavo de pixel, DCT con coeficientes enteros, filtros de deblocking adaptativos y unos sistemas de codificación basados en entropía muy eficientes.

El último avance en el desarrollo de estándares de codificación de vídeo es H.265 [ITU, 2013] o MPEG-H parte 2 (ISO/IEC 23008-2) [ISO, 2013c], denominando comúnmente HEVC (High Efficiency Video Coding). La primera versión del estándar se publicó a principios de 2013 y la segunda ha sido completada en julio de 2014 por lo que se espera su publicación a finales de 2014. HEVC fue diseñado para incrementar de manera notable la eficiencia de codificación en comparación con H.264/AVC High Profile, marcando el objetivo de reducir los requisitos de tasa de bit a la mitad, manteniendo la misma calidad de imagen (a costa de un incremento en la complejidad computacional). HEVC está diseñado para soportar resoluciones de hasta 8192x4320. La estructura de HEVC es similar a la de otros codecs anteriores. Sin embargo, presenta las novedades que se describen a continuación.

- Remplaza la codificación de macrobloques por las unidades de codificación en árbol o CTU (Coding Tree Units), que permiten la codificación conjunta de mayores áreas de la imagen (especialmente conveniente para dar soporte a resoluciones altas).
- Incremento de las direcciones de intra-predicción: a costa de incrementar el tiempo de codificación, HEVC incrementa hasta 35 las posibles direcciones para llevar a cabo predicciones intra-trama, frente a las 9 direcciones utilizadas en H.264.
- Predicción adaptativa de vectores de movimiento, que permite al codificador encontrar más redundancia entre tramas.
- Mejora en las herramientas de paralelización.
- Utilización únicamente de CABAC (Context-Adaptive Binary Arithmetic Coding) como codificador basado en entropía.
- Mejoras en el filtrado de deblocking y un segundo filtrado denominado Sample Adaptive Offset cuyo objetivo es reducir aun más los artefactos en las fronteras entre bloques.

2.5.3. Proceso de codificación

En general, los procesos de codificación que se llevan a cabo para la realización de streaming de vídeo son procesos de compresión con pérdidas, en los que la calidad del vídeo que se obtiene como resultado se puede ver degradada. Como se comentó en la sección anterior, los procesos de codificación que se realizan en los estándares MPEG y H.26x se basan en dos componentes fundamentales: codificación intra-trama usando Discrete Cosine Transform (DCT) y codificación inter-trama utilizando estimación y compensación de movimiento entre tramas de vídeo sucesivas.

En la codificación intra-trama, cada trama se divide en bloques de 8x8 muestras de componentes Y, U y V. Cada bloque se transforma en un bloque de 8x8 coeficientes, utilizando la DCT, que representan las componentes de frecuencia del bloque original. Estos coeficientes se cuantifican mediante una matriz de cuantificación de tamaño 8x8 que contiene los intervalos de cuantificación para cada coeficiente. Este último paso controla el nivel de codificación (y compresión) que obtiene el códec. Por último, se llevan a cabo otros procesos de codificación (escaneo zigzag, codificación run-level y de longitud variable) con el objetivo de reducir todavía más la tasa de bit del códec.

En la codificación inter-trama, se distinguen tres tipos de tramas (ver figura 2.5):

- Tramas I (intra-coded): en una trama I, todos los bloques se codifican mediante codificación intra-trama, como se ha descrito anteriormente.
- Tramas P (inter-coded): en una trama P, cada macrobloque (que consta de 4 bloques de 8x8 muestras), se inter-codifica con respecto a la trama I o P que la precede. Es decir, las tramas I o P anteriores, sirven como referencia para la codificación.
- Tramas B (bidirectional-coded): en una trama B, los macrobloques se inter-codifican con respecto tanto a las tramas I o P que las preceden como a las tramas I o P que las suceden. En las tramas B, se utilizan tramas I o P de referencia tanto anteriores como posteriores.

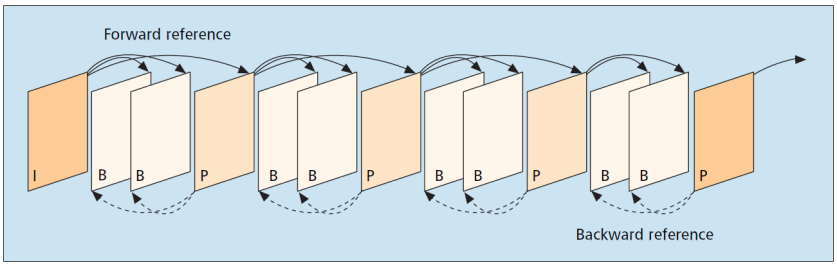


Figura 2.5: Tipos de tramas MPEG

Para inter-codificar un macrobloque se lleva a cabo un proceso de estimación de movimiento con el objetivo de encontrar el macrobloque que mejor se corresponde en la trama de referencia. La diferencia entre el bloque que se quiere codificar y el de referencia se transforma utilizando DCT, se cuantifica y se codifica de manera similar al proceso descrito más arriba. A este proceso se le conoce como compensación de movimiento. En el caso de que no se encuentre ningún bloque de referencia adecuado, el macrobloque se intra-codifica.

La figura 2.5 ofrece una representación gráfica de los distintos tipos de tramas y las relaciones entre ellas. Además, la figura 2.5 corresponde a lo que se conoce como

Group of Pictures (GoP). Un GoP es una secuencia de tramas que se extiende desde una trama I hasta la trama que precede a la siguiente trama I.

2.6. Resumen y conclusiones

En este capítulo se han introducido un conjunto de conceptos generales, aplicables a todo el ámbito de la tesis. Se ha analizado el concepto de calidad, comenzando con definiciones genéricas de calidad, y posteriormente se ha ido centrando el estudio en la calidad de servicio y la calidad percibida, destacando las diferencias y similitudes entre ambos términos.

Se han introducido también una serie de conceptos fundamentales de codificación de vídeo, que servirán como fundamento teórico para el análisis y desarrollo del modelo de calidad de vídeo.

Además, se ha realizado también un análisis de la tecnología MPEG-DASH, principal representante actual del paradigma de la distribución de vídeo adaptativa mediante protocolos fiables, y foco principal de esta tesis. Este tipo de tecnología, supone un cambio en el catálogo de degradaciones de calidad percibida con respecto a otros sistemas, como IPTV, donde la utilización de protocolos no confiables conlleva errores de transmisión que se reflejan en defectos o artefactos en el flujo audiovisual. El streaming de vídeo sobre TCP garantiza que los segmentos de vídeo recibidos están libres de errores, a costa de un retardo de transmisión más elevado que en el caso de UDP. Por tanto, las degradaciones a tener en cuenta para estudiar la calidad percibida en este tipo de servicios son diferentes a las degradaciones de los sistemas clásicos de streaming de vídeo.

Capítulo 3

Estimación de la calidad percibida en servicios de streaming multimedia sobre Internet

3.1. Introducción

El objetivo de este capítulo es desarrollar un modelo que permita estimar la calidad percibida por los usuarios de servicios de streaming multimedia sobre Internet, incluyendo tanto servicios de televisión lineal como de VoD.

Para plantear un modelo global de estimación de calidad percibida es fundamental conocer los componentes del servicio objetivo y entender las relaciones entre dichos componentes. Para ello, este capítulo se apoya en las descripciones de servicios que se realizan en el apéndice A para el servicio de televisión lineal (figura A.2) y VoD (figura A.4). En concreto, el modelo de referencia utilizado se puede ver en la figura 3.1.

A partir de estas descripciones de servicio, en esta sección se presenta el modelo de calidad propuesto, en el cual se combinan las aportaciones a la calidad percibida de las componentes más representativas del servicio.

3.2. Planteamiento general del modelo

El objetivo de este modelo será proporcionar una estimación de la calidad global del servicio, en escala MOS estándar. Dicha estimación deberá ser similar a la que se obtendría al realizar evaluaciones de calidad subjetivas utilizando la escala Absolute Category Rating (ACR) de cinco puntos especificada en ITU-T P.800 [ITU, 1997b],

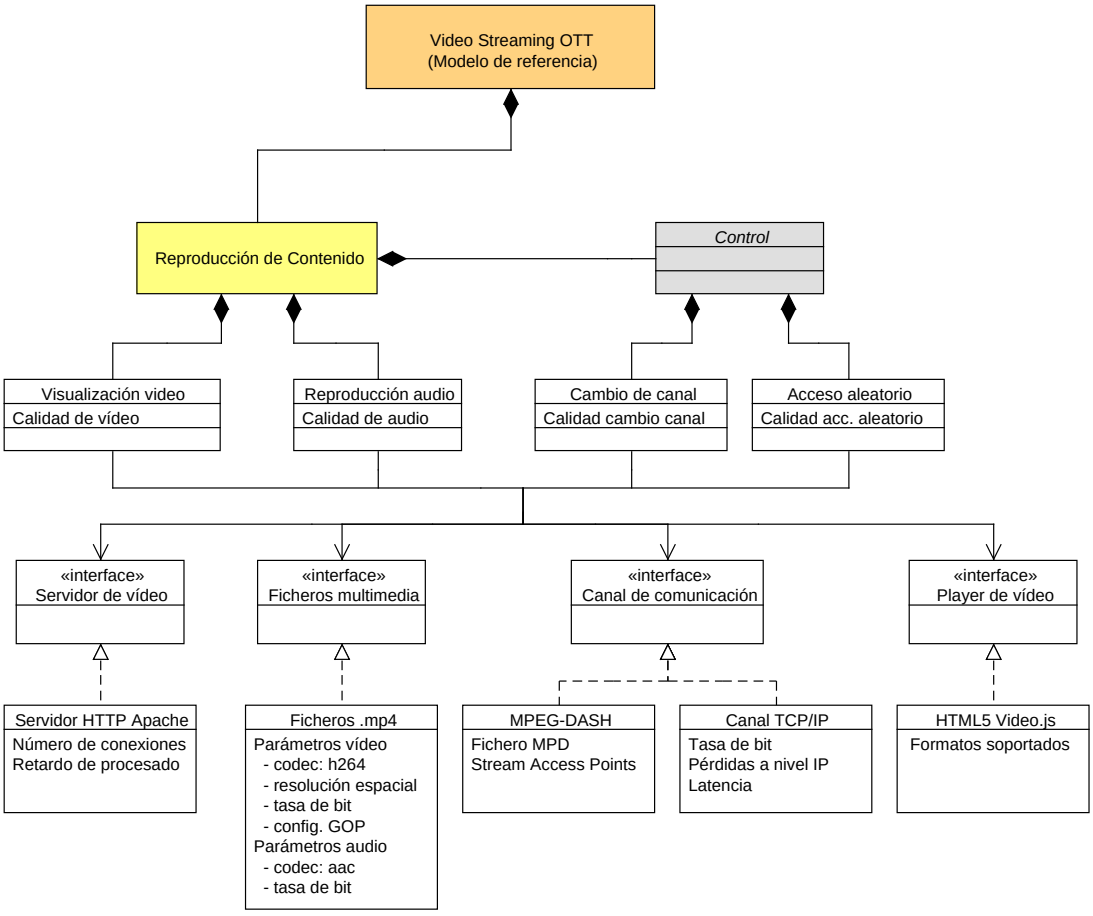


Figura 3.1: Modelo de referencia del servicio de streaming multimedia sobre Internet

ITU-T P.800.1 [ITU, 2006] e ITU-T P.910 [ITU, 2008f] para Puntuación de Opinión Media (MOS).

Las aportaciones de calidad que se consideran para el desarrollo del modelo global de calidad son las siguientes:

- Calidad de vídeo
- Calidad de audio
- Calidad (o degradación) asociada a la sincronización entre el audio y el vídeo (en inglés, lip-sync)
- Degradación asociada al efecto de la red (degradación por transmisión).
- Tiempo de seeking o acceso aleatorio
- Tiempo de cambio de canal

En general, en la literatura de esta área de conocimiento no hay modelos bien establecidos que permitan la estimación de la calidad global del servicio de streaming de vídeo adaptativo OTT a partir de estas componentes. En [de la Cruz Ramos, 2012] se lleva a cabo una revisión del estado del arte con un objetivo similar, para el caso del servicio de difusión de televisión IPTV, y tras poner de manifiesto la falta de modelos globales de este tipo, propone tres tipos de modelos con los que combinar las distintas aportaciones de calidad de cada una de las componentes del servicio:

- Modelo lineal
- Modelo no lineal
- Variaciones de los anteriores, utilizando factores de degradación (I_x) en vez de valoraciones de calidad (Q_x).

En esta tesis, se ha optado por la utilización de un modelo lineal, similar al utilizado en [de la Cruz Ramos, 2012], con ciertas variantes que se describen en las siguientes secciones. En dicha descripción se ha optado por utilizar un enfoque "top-down", por lo que se comienza describiendo el modelo desde un punto de vista de alto nivel para después ir añadiendo detalles de cada una de las partes del mismo.

3.2.1. Escalas de calidad y nomenclatura

Antes de comenzar con la descripción del modelo, conviene concretar algunos aspectos relacionados con las distintas escalas de calidad y la nomenclatura utilizadas a lo largo del capítulo.

Dentro del campo de estudio de la calidad percibida en servicios de telecomunicación, se han desarrollado y aplicado diferentes escalas de calidad a diferentes tipos de tests subjetivos de calidad. Cada una de estas escalas viene definida por su carácter discreto (escalas categóricas) o continuo (escalas gráficas), el número de niveles de calidad y la semántica asociada a cada nivel. Algunas de las escalas de calidad más utilizadas son las siguientes:

- Escala discreta de 5 puntos, incluida en ITU-T P.910 [ITU, 2008f] y ITU-R BT.500 [ITU, 2012a] y ampliamente usada en la literatura. El significado asociado a cada nivel de calidad es el siguiente: {5: Excelent, 4: Good, 3: Fair, 2: Poor, 1: Bad}. Además, apoyándose en esta escala, la industria ha adoptado el concepto de MOS como un valor entre 1 y 5.
- Escala discreta de 9 puntos, incluida en el anexo de la recomendación ITU-T P.910 [ITU, 2008f] como una escala especialmente indicada para la evaluación de calidad de codecs de vídeo de baja tasa de bit. Dicha escala se basa en incluir un

nivel intermedio entre cada nivel de la escala discreta de 5 puntos, por lo que el significado asociado a cada nivel de calidad es el similar: {9: Excelent, 7: Good, 5: Fair, 3: Poor, 1: Bad}.

- Escala discreta de 11 puntos, incluida en el anexo de la recomendación ITU-T P.910 [ITU, 2008f], como una extensión a la escala de 9 puntos en la que se han añadido dos niveles (uno superior, el nivel 10; y otro inferior, el nivel 0). El nivel 10 representa un nivel de calidad que no admite mejora y el nivel 0 un nivel de calidad tal que no puede imaginarse una calidad peor.

- A partir de las dos escalas anteriores, la recomendación ITU-T P.910 [ITU, 2008f] define dos escalas de carácter continuo:

  - Escala continua de 9 puntos
  - Escala continua de 11 puntos

Un aspecto interesante, estudiado en [Huynh-Thu et al., 2011], es la relación entre las diferentes escalas. En dicho estudio remarcan las siguientes ideas:

- La mayor parte de los participantes en tests de evaluación de calidad tienden a alinear sus valoraciones con las etiquetas de las escalas (en el caso de escalas continuas).

- Existe una fuerte relación lineal y no hay diferencias estadísticas significativas entre los resultados obtenidos utilizando diferentes escalas.

- Los tests subjetivos basados en la presentación de un solo estímulo (como el método ACR), diseñados adecuadamente e informando correctamente a los participantes, producen resultados repetibles incluso utilizando diferentes escalas.

Así pues, a lo largo de esta tesis, **las valoraciones individuales de calidad se expresarán mediante una escala discreta de 5 puntos. Por tanto, la media de valoraciones de calidad entre individuos o MOS será expresada como un valor continuo en el rango [1, 5]**.

Además de esta escala, en ciertos modelos se utiliza una escala cuyo rango es [0, 100]. Dicha escala está inspirada en el factor R del modelo-E (definido en ITU-T G.107 [ITU, 2014b]), que asume que distintas degradaciones expresadas en esta escala son aditivas.

En esta tesis, **teniendo en cuenta el carácter aditivo en escala R, la mayor parte de los modelos se desarrollarán utilizando esta escala, transformando finalmente el resultado global a un valor de MOS mediante la expresión correspondiente**. A continuación se analizan distintas expresiones para transformar valoraciones de calidad entre escala MOS y escala R y viceversa.

La recomendación ITU-T G.107 [ITU, 2014b] ofrece una expresión para realizar la transformación entre la escala R y la escala MOS.

$$MOS(R) = \begin{cases} 4,5, & Q \geq 100 \\ 1 + 0,035 \cdot R + R \cdot (R - 60) \cdot (100 - R) \cdot 7 \cdot 10^{-6}, & 0 < Q < 100 \\ 1, & R \leq 0 \end{cases} \quad (3.1)$$

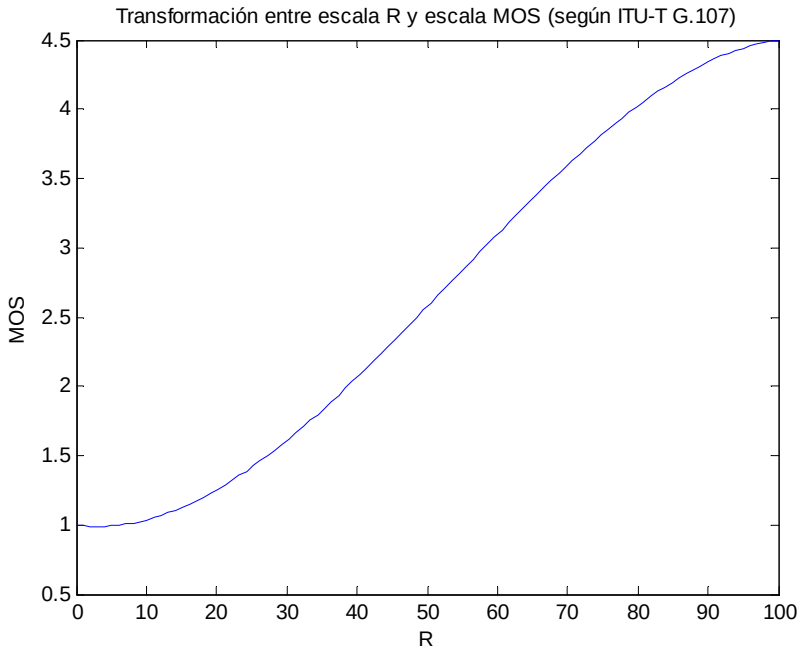


Figura 3.2: Relación entre escala R y escala MOS según ITU-T G.107

Por otro lado, la recomendación ITU-T P.1201.2 [ITU, 2012e] (de especial relevancia para esta tesis, como se verá más adelante) propone algunas variaciones con respecto a la transformación de escalas proporcionada por el modelo-E.

$$MOS(R) = \begin{cases} 4,9, & Q \geq 100 \\ 1 + 0,0385 \cdot R + R \cdot (R - 60) \cdot (100 - R) \cdot 7 \cdot 10^{-6}, & 0 < Q < 100 \\ 1,05, & R \leq 0 \end{cases} \quad (3.2)$$

Como se puede ver en la ecuación 3.2, en esta expresión se ha aumentado tanto el valor máximo como el valor mínimo de MOS que se puede obtener a 4,9 y 1,05

respectivamente (en la transformación del modelo-E, dichos valores son 4,5 y 1 respectivamente). Sin embargo, la expresión anterior tiene un problema de continuidad, que puede ser debido a una errata en la recomendación (existen enmiendas de la misma, pero no son accesibles al público). Si se evalúa la expresión anterior en los puntos extremos de la misma (R=0 y R=100), se puede ver que los valores a ambos lados de dichos puntos no coinciden. Este problema puede expresarse de manera más formal mediante la ecuación 3.3.

$$\begin{aligned} \lim_{R\to 100^-} MOS(R) &\neq \lim_{R\to 100^+} MOS(R) \\ \lim_{R\to 0^-} MOS(R) &\neq \lim_{R\to 0^+} MOS(R) \end{aligned} \tag{3.3}$$

Este "problema" se puede solucionar modificando la expresión como se muestra en la ecuación 3.4.

$$MOS(R) = \begin{cases} 4,9, & R \geq 100 \\ 1,05 + 0,0385 \cdot R + R \cdot (R-60) \cdot (100-R) \cdot 7 \cdot 10^{-6}, & 0 < R < 100 \\ 1,05, & R \leq 0 \end{cases} \tag{3.4}$$

Debido a que el ámbito de la recomendación ITU-T P.1201.2 es más afín a esta tesis que el modelo-E, **la ecuación 3.4 será la que se utilice en el resto de la tesis para realizar conversiones entre escala R y escala MOS**.

Por último, se debe destacar que, a diferencia de la recomendación ITU-T G.107, que proporciona una expresión para realizar la conversión de escala MOS a escala R, la recomendación ITU-T P.1201.2 no proporciona dicha expresión, por lo que se ha tenido que desarrollar una nueva.

El proceso seguido para el desarrollo de esta función (inversa de la función anterior $MOS(R)$) ha sido partir de la representación gráfica de la función propuesta en ITU-T G.107 y adaptar los valores extremos de la misma al nuevo rango de valores utilizado. Así pues, se han forzado las siguientes condiciones:

$$\begin{aligned} R(MOS = 1,05) &\approx 0 \\ R(MOS = 4,9) &\approx 100 \end{aligned} \tag{3.5}$$

Además, viendo la similitud de la función propuesta en ITU-T G.107 con la función logit, se ha optado por no utilizar directamente la compleja expresión de ITU-T G.107, sino realizar un nuevo ajuste numérico sobre una variación de la función logit. Así pues,

obviando los detalles de dicho ajuste, **la expresión que se va a utilizar en el resto de la tesis para transformar un valor en escala MOS a un valor en escala R se muestra en la ecuación 3.6**.

$$R(MOS) = c_1 \cdot log\left(\frac{MOS - 1}{c_2 - (MOS - 1)}\right) + c_3 \cdot e^{-c_4 \cdot MOS} + c_5 \tag{3.6}$$

Los parámetros de ajuste para la ecuación 3.6 se presentan en la tabla 3.1:

Tabla 3.1: Parámetros de ajuste para la función de conversión entre escala R y escala MOS

Parámetros				
c_1	c_2	c_3	c_4	c_5
19,1	4,005	76,44	-0,08503	-46,31

Con estos parámetros de ajuste, la curva que se obtiene se puede ver en la figura 3.3.

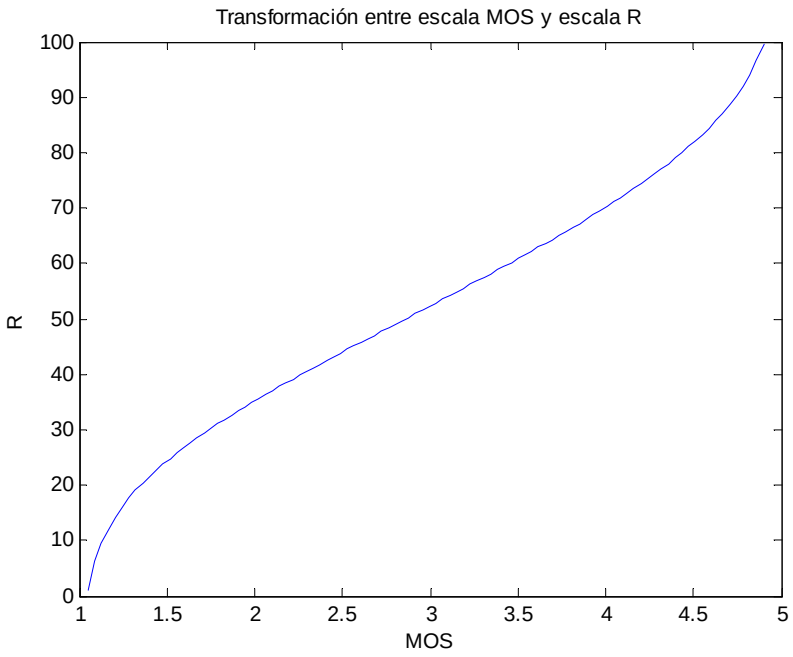


Figura 3.3: Relación propuesta entre escala MOS y escala R

3.3. Modelo global de estimación de QoE de un servicio de streaming de vídeo a partir de las valoraciones de calidad de sus componentes

Como se introdujo en las secciones anteriores, el modelo que se propone en esta tesis tiene como objetivo proporcionar una estimación de la calidad global de un servicio de streaming mediante la combinación lineal de un conjunto de valoraciones de calidad asociadas a cada componente del servicio.

Un aspecto importante a tener en cuenta es que no todos los componentes de servicio actúan o tienen efecto durante todo el periodo temporal de prestación del servicio. En este contexto, los distintos componentes de un servicio que afectan a la calidad percibida pueden ser clasificados en función de la fracción del tiempo de prestación del servicio en la que dicho componente está presente o tiene efecto para el usuario. Así pues, en esta tesis la hipótesis que se plantea es que el efecto de los componentes que se van a denominar **"componentes continuos"** en la calidad total es distinto a los **"componentes puntuales"**, entendiendo como componentes continuos aquellos cuyo efecto está presente durante la mayor parte del tiempo de prestación del servicio, frente a componentes puntuales, que son aquellos cuyo efecto solo aplica en intervalos de tiempo limitados. En el caso de un servicio de streaming de vídeo adaptativo OTT, componentes continuos serían los componentes de servicio como "Visualización de vídeo", "Reproducción de audio", mientras que los componentes "Cambio de canal" o "Acceso aleatorio" serían componentes puntuales.

Así pues, una vez introducidos los dos tipos de componentes, se podría expresar la calidad global de un servicio según la ecuación 3.7, donde Q es la estimación de calidad global, Q_C es el factor de calidad de todos los componentes continuos, N_c es el número de componentes continuos, c_i es el peso de cada componente continuo, Qc_i es el factor de calidad del componente continuo i-ésimo, N_p es el número de componentes puntuales, p_j es el peso de cada componente puntual y Qp_j es el factor de calidad del componente puntual j-ésimo.

$$Q = \sum_{i=1}^{N_c} c_i \cdot Q_{c_i} + \sum_{j=1}^{N_p} p_j \cdot Q_{p_j} = Q_C + \sum_{j=1}^{N_p} p_j \cdot Q_{p_j} \tag{3.7}$$

Aunque la ecuación 3.7 está escrita en términos de factores de calidad, a lo largo de este capítulo se verá que es más conveniente reescribir algunos de estos términos en forma de factores de degradación, aplicando $Q_x = Q_{max} - I_x$, siendo I_x el factor de degradación asociado al componente de servicio x.

A continuación se detalla cada uno de los componentes para el caso del servicio considerado en esta tesis.

3.3.1. Componentes continuos

Como se dijo anteriormente, se definen como componentes continuos aquellos componentes de un servicio cuyo efecto está presente durante todo o prácticamente todo el tiempo de prestación de dicho servicio. En el caso del servicio de streaming de vídeo adaptativo OTT, los factores de calidad asociados a componentes continuos considerados para el modelo global de calidad son los siguientes:

- Calidad audiovisual: esta componente engloba las valoraciones de calidad de la componente de audio y video y su interacción (ver 3.4.1).
- Calidad (o degradación) asociada al lipsync: esta componente contempla el efecto que tiene la (de)sincronización entre los flujos de audio y vídeo (ver 3.4.2).
- Calidad (o degradación) asociada a la transmisión: esta componente contempla el efecto que tiene la red en la calidad percibida (ver 3.4.3).

En la literatura de esta área, la calidad audiovisual engloba las valoraciones de calidad de los flujos de audio y vídeo, suponiendo generalmente que dichos flujos están sincronizados. Para cuantificar el efecto que tiene la falta de sincronización entre flujos es común añadir otro factor de calidad al modelo.

Por ejemplo, en [de la Cruz Ramos, 2012], la aportación a la calidad global de las componentes audiovisual y de sincronización audio-vídeo se modela según la ecuación 3.8, donde $Q_{av_{total}}$, Q_{av} y Q_{ls} representan la calidad audiovisual total, la calidad audiovisual (suponiendo sincronización entre los flujos de audio y vídeo) y la calidad asociada a la sincronización entre flujos de audio y vídeo, respectivamente. Los factores c_{av} y c_{ls} modelan la importancia de cada una de las componentes de calidad que constituyen la calidad audiovisual total.

$$Q_{av_{total}} = c_{av} \cdot Q_{av} + c_{ls} \cdot Q_{ls} = 0,75 \cdot Q_{av} + 0,18 \cdot Q_{ls} \tag{3.8}$$

Sin embargo, en esta tesis los factores de calidad audiovisual y de sincronización audio-vídeo se van a combinar de manera diferente. Considérense las siguientes hipótesis:

- Si Q_{av} es muy baja, entonces $Q_{av_{total}}$ debería ser baja (independientemente del nivel de sincronización entre flujos).
- Si la sincronización entre audio y vídeo es muy mala, entonces $Q_{av_{total}}$ debería ser baja (independientemente del valor de Q_{av})
- El efecto del lipsync depende del tipo de contenido (ejemplo: un noticiario frente a un partido de fútbol, en el noticiario el efecto de la sincronización es mayor que en el partido de fútbol). Ver sección 3.4.2.

Si se vuelve a analizar la ecuación 3.8 [de la Cruz Ramos, 2012], se puede ver que no se respetan las hipótesis anteriores. Por ejemplo, si $Q_{av} = 100$ (máxima calidad audiovisual) y $Q_{ls} = 0$ (mínima calidad asociada al lip-sync), entonces $Q_{av_{total}} = 75$, valor relativamente alto que no modela adecuadamente una secuencia de vídeo cuya calidad audiovisual es muy buena, pero en la que la sincronización entre los flujos de audio y vídeo es muy deficiente. Este tipo de secuencias, dependiendo del tipo de contenido, pueden ser evaluadas muy negativamente por un usuario, debido a la dificultad que entraña el visualizar un contenido en el que existe tanta desincronización entre los flujos de audio y vídeo.

Si se tienen en cuenta estas hipótesis, no se puede plantear el efecto de la sincronización audio-vídeo como un factor de calidad, sino como un factor de degradación. Expresado en términos matemáticos, el efecto del lipsync debería ser una cantidad negativa o nula. Como Q_{av} presupone sincronización perfecta entre audio y vídeo, todo efecto que provenga de una mala sincronización entre audio y vídeo deberá perjudicar (restar) a la calidad total.

Así pues, se propone la ecuación 3.9 para estimar la calidad audiovisual total de un servicio, siendo $Q_{av_{total}}$ la calidad audiovisual total, Q_{av} el factor de calidad audiovisual suponiendo sincronización perfecta entre audio y vídeo e I_{ls} un factor de degradación que cuantifica el efecto que tiene la desincronización entre los flujos de audio y vídeo.

$$Q_{av_{total}} = Q_{av} - I_{ls} \tag{3.9}$$

Una vez definida la expresión para la estimación de la calidad audiovisual total, a continuación se discute cómo se incluye en el modelo el efecto que tiene la transmisión de los flujos de vídeo por la red.

Como se puede ver en el modelo de referencia del servicio contemplado (figura 3.1), la transmisión del contenido se realiza a través una red TCP/IP no gestionada (vídeo OTT) mediante MPEG-DASH. El hecho de utilizar un protocolo de transporte fiable conlleva que los fragmentos de vídeo MPEG-DASH que envía el cliente son recibidos sin errores y con un cierto retardo. Este retardo es el principal causante de las posibles degradaciones en la calidad que pueden producirse en un servicio de vídeo OTT, dando lugar a tiempos de espera e interrupciones en la reproducción del contenido, además de variaciones en el nivel de calidad de vídeo a lo largo del tiempo (ver capítulo 5).

Así pues, la red puede introducir un conjunto de degradaciones que tienen un efecto negativo sobre la calidad audiovisual total, lo cual se puede modelar mediante la ecuación 3.10.

$$Q_C = Q_{av_{total}} - I_{tra} \tag{3.10}$$

3.3.2. Componentes puntuales

Los factores de calidad asociados a componentes puntuales que van a ser considerados en el modelo de calidad de esta tesis son los siguientes:

- Degradación de calidad asociada al cambio de canal: este factor de calidad estima el efecto que tiene en la calidad percibida el tiempo que necesita el servicio para llevar a cabo un cambio de canal.
- Degradación de calidad asociada al acceso aleatorio: este factor de calidad estima el efecto que tiene en la calidad percibida el tiempo y la precisión en el acceso aleatorio a un punto arbitrario de la línea temporal del contenido.

Como se puede ver, se ha planteado el efecto de los componentes puntuales como una degradación. En el caso ideal, tanto el cambio de canal como el acceso aleatorio se realizarían de manera instantánea, por lo que si un servicio requiere una cantidad de tiempo no despreciable para realizar estas operaciones, ésto conllevará un decremento en la calidad percibida.

Así pues, la ecuación 3.7 quedaría modificada como sigue:

$$Q = Q_C - \sum_{j=1}^{N_p} p_j \cdot I_{p_j} \tag{3.11}$$

De manera análoga al razonamiento seguido para modelar la interacción entre la calidad audiovisual y el efecto de la sincronización entre flujos, a continuación se enumeran varias hipótesis que se van a utilizar para modelar la contribución de los componentes puntuales a la calidad global del servicio.

- La influencia de la calidad asociada a los componentes puntuales I_{p_j} es relevante para el cómputo de la calidad total Q solo si la calidad de la totalidad de los componentes continuos Q_C alcanza un cierto valor: si la calidad de los componentes continuos es baja, el nivel de calidad de los componentes puntuales es poco relevante. Por ejemplo, un servicio de TV con calidad audiovisual baja, será percibido como un servicio de mala calidad, independientemente de lo buenos o malos (en términos de calidad percibida) que sean otros factores como el cambio de canal.
- La influencia de I_{p_j} puede ser moderada si Q_C supera un cierto valor: si la calidad de los componentes continuos es muy alta, la tolerancia en cuanto a la calidad de los componentes puntuales (que afectan durante una fracción de tiempo pequeña) puede ser mayor, es decir, su relevancia puede verse moderada. Por ejemplo: en un servicio de TV con una calidad audiovisual muy alta, el que el nivel de calidad del cambio de canal sea moderado, no influirá demasiado en la valoración de la calidad total.

Teniendo en cuenta estas hipótesis, lo que se propone es modelar el efecto de los componentes puntuales teniendo en cuenta la dependencia de dicho efecto con la calidad de los componentes continuos. Más concretamente, se propone la siguiente definición, en la que a los factores p_j se les ha añadido una componente que es función de Q_C:

$$p_j = p_{j_0} \cdot w_j = p_{j_0} \cdot f(Q_C) \tag{3.12}$$

Teniendo en cuenta las hipótesis anteriores, se propone la utilización de una función como la que se presenta en la figura 3.4. Se debe destacar que la forma de curva propuesta es una aproximación basada en las hipótesis anteriores, por lo que la obtención de una curva más exacta, como resultado de experimentos de valoración subjetiva de calidad, se deja como trabajo futuro.

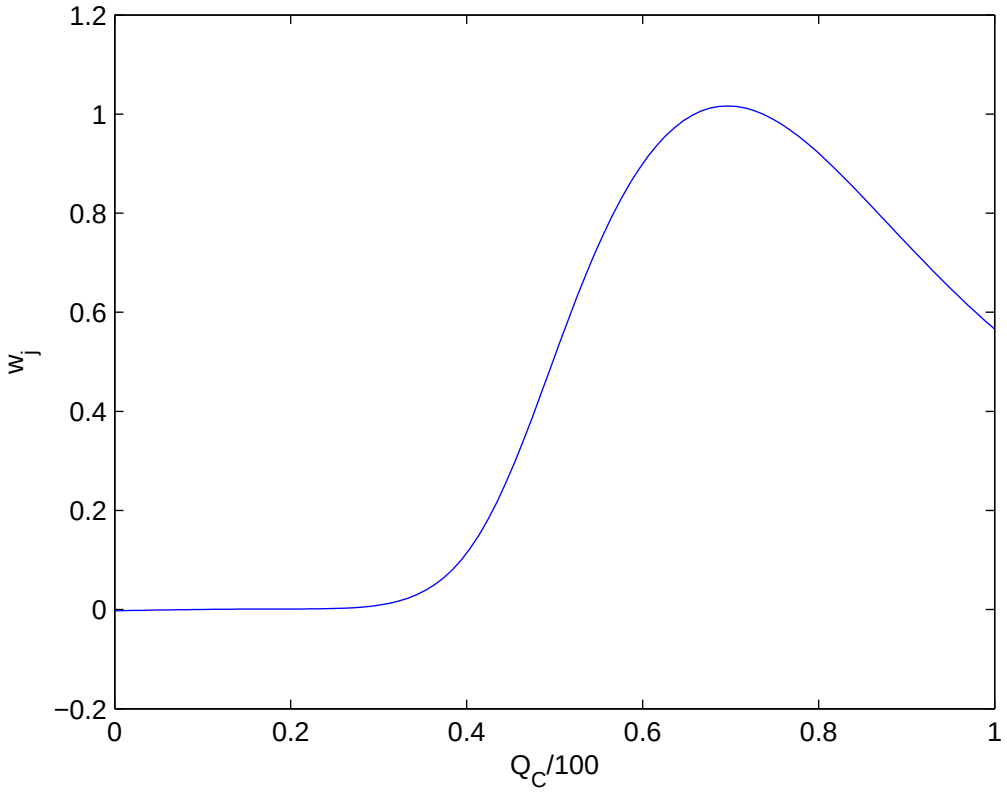


Figura 3.4: Función $f(Q_C)$ propuesta

Esta forma de curva puede ser ajustada numéricamente mediante la ecuación 3.13.

$$f(x) = \frac{r_3 \cdot x^3 + r_2 \cdot x^2 + r_1 \cdot x + r_0}{x^4 + q_3 \cdot x^3 + q_2 \cdot x^2 + q_1 \cdot x + q_0} \tag{3.13}$$

Los parámetros de ajuste de $f(Q_C)$ se presentan en la tabla 3.2.

Tabla 3.2: Parámetros de ajuste para la función $f(Q_C)$

Parámetros							
r_3	r_2	r_1	r_0	q_3	q_2	q_1	q_0
0,06974	-0,03868	0,006936	-0,0003597	-2,293	2,085	-0,8652	0,1397

En las siguientes secciones se analiza en detalle cada uno de los factores del modelo de calidad introducidos hasta el momento.

3.4. Componentes continuos

3.4.1. Estimación del factor de calidad audiovisual para flujos sincronizados

La calidad de audio y vídeo son campos en los que se ha llevado a cabo una extensa labor de investigación, tanto a nivel de evaluación subjetiva, como a nivel de estimación objetiva de calidad. Sin embargo, la calidad audiovisual es un área en la que no se han llevado a cabo tanto esfuerzos.

El resto de la sección se organiza de la siguiente manera: en primer lugar se realiza una revisión de la literatura relacionada con la estimación de la calidad audiovisual para flujos sincronizados y tras ello, se introduce el modelo propuesto en esta tesis, el cual combina soluciones existentes en la literatura con aportaciones propias.

3.4.1.1. Revisión del estado del arte

En [Château, 1998] se analiza la influencia entre la calidad de audio y vídeo en contextos de videoconferencia, llegando a la conclusión de que la calidad audiovisual depende fuertemente de la calidad de vídeo, mientras que la calidad de audio tiene un efecto más débil, pero no despreciable, en la calidad audiovisual total. Además, afirman que la calidad del vídeo influencia la calidad percibida del audio, mientras que la percepción de la calidad del vídeo es independiente de la calidad del audio.

En [Beerends and De Caluwe, 1999] se lleva a cabo un estudio de la calidad audiovisual en aplicaciones de videoconferencia mediante la simulación de distorsiones analógicas. Dicho estudio pone de manifiesto que la calidad multimedia depende del tipo de contenido de las secuencias audiovisuales. Más concretamente, los autores afirman que para secuencias con poca información temporal (secuencias de tipo busto parlante), tanto el audio como el vídeo tienen un efecto significativo en la percepción de la calidad. Sin embargo, en secuencias con más movimiento, la calidad del vídeo tiene una mayor contribución a la calidad total. En cuanto a la influencia entre la calidad del audio y del

vídeo, este estudio afirma que la calidad de audio y vídeo tienen una cierta influencia mutua: la calidad del vídeo tiene una influencia relativamente fuerte en la percepción de la calidad del audio (del orden del 13 %) mientras que la influencia de la calidad de audio sobre la calidad percibida del vídeo es más débil (del orden del 2 %). En [ITU, 1997a] [ITU, 1998a] se llega a conclusiones similares.

En [Joly et al., 2001] se estudia la calidad audiovisual en televisión digital, llegando a la conclusión de que la calidad del vídeo no tiene influencia sobre la calidad del audio, pero la calidad de audio sí que tiene influencia sobre la percepción de las degradaciones del vídeo. Además, afirma que la calidad audiovisual depende fuertemente de la calidad de vídeo, mientras que la contribución de la calidad de audio es más débil, pero no despreciable.

En [Pastrana-Vidal et al., 2003] se lleva a cabo una revisión de diversos modelos de calidad audiovisual, todos ellos casos particulares de un modelo general que se verá más adelante.

La recomendación ITU-T J.148 [ITU, 2003] especifica los requisitos que debe cumplir un modelo objetivo de calidad percibida en servicios multimedia. La arquitectura recomendada se muestra en la figura 3.5.

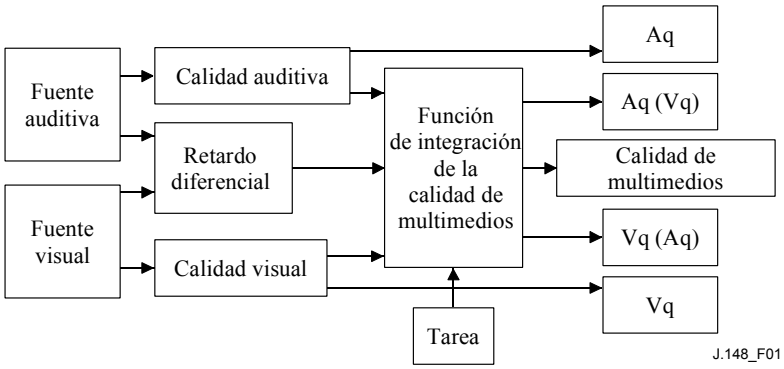


Figura 3.5: Arquitectura de un modelo de calidad multimedia según ITU-T J.148

Como se puede ver en la figura 3.5, a partir de las fuentes de audio y vídeo, se extraen tanto la calidad visual y auditiva, como el retardo relativo entre las fuentes de audio y vídeo. Una vez obtenidos estos valores, se utilizan como entradas para un módulo cuya misión es integrar los distintos valores de calidad. Dicho módulo cuenta con una entrada adicional que permite incluir aspectos dependientes de la tarea (grado de interactividad, etc.).

En [Hands, 2004] se propone un modelo de predicción de calidad audiovisual a partir de modelos de calidad de audio y vídeo que estiman la calidad percibida del audio y del vídeo de manera independiente. Con el objetivo de derivar (mediante un análisis de

regresión) las reglas necesarias para integrar estas dos estimaciones de calidad, el autor llevó a cabo dos experimentos: el primero de ellos basado en contenido de tipo busto parlante y el segundo basado en una combinación de contenidos de tipo busto parlante y contenido con mucho movimiento.

El modelo extraído del primer experimento se muestra en la ecuación 3.14. Como se puede ver, al tratarse de contenido con muy poco movimiento, la calidad del audio tiene mayor influencia que la calidad del vídeo.

$$Q_{multimedia} = 0,85 \cdot Q_{audio} + 0,76 \cdot Q_{video} - 0,01(Q_{audio} \cdot Q_{video}) - 3,34 \tag{3.14}$$

El modelo extraído del segundo experimento, para contenidos con mucho movimiento, se muestra en la ecuación 3.15, poniendo de manifiesto que en contenidos con mucho movimiento la calidad del vídeo tiene una mayor influencia que la calidad del audio.

$$Q_{multimedia} = 0,25 \cdot Q_{video} + 0,15 \cdot (Q_{audio} \cdot Q_{video}) + 0,95 \tag{3.15}$$

Así pues, en general las conclusiones que se extraen de estos modelos son las siguientes:

- El tipo de contenido de la secuencia influye en la percepción de la calidad audiovisual.
- El término multiplicativo audio-vídeo no es despreciable.

En [Winkler and Faller, 2005] y en [Winkler and Faller, 2006] se llevan a cabo una serie de experimentos subjetivos de calidad de audio, vídeo y audiovisual, mediante contenidos representativos de aplicaciones móviles, con los objetivos de, en primer lugar, analizar las interacciones entre el audio y el vídeo en términos de calidad audiovisual, y en segundo lugar, evaluar el rendimiento de una serie de modelos de calidad de audio y vídeo sin referencia para estimar la calidad audiovisual.

Para el análisis de la influencia de la calidad de audio, de la calidad de vídeo y del término de interacción multiplicativo en la calidad audiovisual, los autores realizaron un análisis de componentes principales o Principal Component Analysis (PCA). Como resultado de este análisis, y apoyándose en los resultados de [Hands, 2004], los autores proponen dos modelos, el primero de ellos multiplicativo y el segundo lineal, como se puede ver en la ecuación 3.16.

$$\begin{aligned} MOS_{av} &= 1,98 + 0,103 \cdot MOS_a \cdot MOS_v \\ MOS_{av} &= -1,51 + 0,456 \cdot MOS_a + 0,77 \cdot MOS_v \end{aligned} \tag{3.16}$$

Como se puede ver, los modelos propuestos confirman las ideas propuestas por otros autores en cuanto a que la calidad del vídeo tiene una mayor influencia en la calidad audiovisual que la calidad del audio.

En la recomendación ITU-T G.1070 [ITU, 2012c] se recomienda un modelo de estimación de la Calidad Audiovisual Percibida para aplicaciones de videotelefonía interactiva punto-a-punto sobre redes IP. El modelo consta de tres funciones: estimación de la calidad de vídeo, estimación de la calidad de voz (audio) y función de integración de la calidad multimedia.

Para estimar la calidad multimedia, el modelo incluye una fase intermedia con el objetivo de estimar la calidad audiovisual. Esta estimación intermedia se presenta en la ecuación 3.17, donde MM_{SV} es la calidad audiovisual, S_q es la calidad de la voz, V_q es la calidad del vídeo y m_i son coeficientes que dependen del tamaño de la imagen y de la tarea conversacional específica.

$$MM_{SV} = m_5 S_q + m_6 V_q + m_7 S_q V_q + m_8 \tag{3.17}$$

A partir de la expresión de la calidad audiovisual y de un factor de degradación asociado a la desincronización de los flujos de audio y vídeo, la recomendación ITU-T G.1070 proporciona un modelo para estimar la calidad multimedia total. Este modelo se presenta en la ecuación 3.18, donde MM_q es la calidad multimedia total, MM_{SV} es la calidad audiovisual (ecuación 3.17), MM_T representa la degradación debida a la desincronización de los flujos de audio y vídeo y m_i son coeficientes que dependen del tamaño de imagen y de la tarea conversacional específica.

$$MM_q = m_1 MM_{SV} + m_2 MM_T + m_3 MM_{SV} MM_T + m_4 \tag{3.18}$$

El factor de degradación de puede estimar utilizando las ecuaciones 3.19 3.20 y 3.21, donde AD es el retardo audiovisual absoluto, MS es el factor de sincronización audiovisual, T_S es el retardo de voz extremo a extremo en un sentido, T_V es el retardo de vídeo extremo a extremo en un sentido y m_i son coeficientes que dependen del tamaño de imagen y de la tarea conversacional específica.

$$MM_T = max\{AD + MS, 1\} \tag{3.19}$$

$$AD = m_9 \cdot (T_S + T_V) + m_{10} \tag{3.20}$$

$$MS = \begin{cases} min\{m_{11} \cdot (T_S - T_V) + m_{12}, 0\}, & T_S \geq T_V \\ min\{m_{13} \cdot (T_S - T_V) + m_{14}, 0\}, & T_S < T_V \end{cases} \tag{3.21}$$

En la tabla 3.3 se incluye un conjunto de valores para los coeficientes m_i que proporciona la recomendación.

Tabla 3.3: Coeficientes del modelo ITU-T G.1070

Parámetros	**Tamaño de la pantalla**	
	2,1”	4,2”
m_1	$-0,6966$	$-0,4457$
m_2	$-0,8127$	$-0,6638$
m_3	0,4562	0,4042
m_4	3,003	2,321
m_5	$-0,1638$	$-0,3255$
m_6	0,3626	0,3309
m_7	1,291	1,494
m_8	0,5456	0,5457
m_9	$-1,251 \cdot 10^{-4}$	$-3,235 \cdot 10^{-4}$
m_{10}	3,763	3,915
m_{11}	$-1,065 \cdot 10^{-3}$	$-1,377 \cdot 10^{-3}$
m_{12}	$1,465 \cdot 10^{-2}$	0
m_{13}	$-1,002 \cdot 10^{-3}$	$-1,095 \cdot 10^{-3}$
m_{14}	0	0

En [Winkler and Mohandas, 2008] se identifican los principales factores que influyen en la calidad audiovisual:

- La calidad de audio
- La calidad de video
- La interacción entre la calidad de audio y de video
- La sincronización entre el audio y el vídeo

Además, menciona otros factores, que típicamente no se tienen en cuenta en los modelos propuestos en la literatura, pero cuya influencia no es para nada despreciable en la estimación de la calidad percibida:

- Nivel de atención o de interés del usuario que visualiza el contenido
- Expectativas del usuario
- Experiencia del usuario en cuanto a servicios o tecnología de vídeo, la cual determina o influye en las expectativas del mismo
- Tipo de pantalla
- Condiciones de visionado

En [Maki et al., 2013] se presenta un modelo paramétrico de referencia reducida para la estimación de la calidad audiovisual en IPTV y servicios similares. Este modelo extrae ciertas características de la secuencia original, relacionadas con el nivel de movimiento del contenido. Más concretamente, utilizan una métrica definida como MQ-C, que se calcula como la suma de los valores de Spatial Information (SI) y Temporal Information (TI) de la secuencia original. Además, al ser un modelo paramétrico, utiliza otro conjunto de datos que se puede extraer de las cabeceras de los paquetes. Estos parámetros son los siguientes:

- Resolución
- Porcentaje de pérdidas de paquete
- MLBS (tamaño medio de la ráfaga de pérdidas)

Con estos inputs, los autores han entrenado una red neuronal (basada en una arquitectura de perceptrón multicapa) para obtener la estimación de la calidad audiovisual.

En [Garcia and Raake, 2009] se propone un modelo de estimación de calidad audiovisual para servicios de IPTV. Dicho modelo considera las degradaciones introducidas tanto en el proceso de compresión del audio y el vídeo como en la transmisión (errores de paquete). Analizan dos versiones del modelo, una de ellas basada en factores de degradación y otra basada en factores de calidad, mostrando ligeramente mejores resultados la primera de ellas. En general, los resultados demuestran la influencia mutua entre la calidad percibida de los flujos de audio y de vídeo y la predominancia de la calidad de vídeo en la valoración de calidad audiovisual (ver figura 3.6).

Como se puede ver en la figura 3.6, la influencia de cada uno de los flujos en la calidad audiovisual depende de la calidad del otro flujo: por ejemplo, la calidad de audio tiene una influencia decreciente en la calidad audiovisual conforme decrece la calidad de vídeo.

El modelo de calidad audiovisual de partida que utilizan en el desarrollo de su trabajo se muestra en la ecuación 3.22.

$$Q_{av} = \alpha + \beta \cdot Q_a + \gamma \cdot Q_v + \mu \cdot Q_a \cdot Q_v \tag{3.22}$$

Además, realizan una transformación de este modelo general para utilizar factores de degradación en vez de factores de calidad, como se puede ver en la ecuación 3.23, donde $[a..i]$ son parámetros de ajuste del modelo, $Icod_x$ es el factor de degradación asociado al proceso de codificación o compresión del flujo de vídeo ($x = V$) o audio ($x = A$) e $Itra_x$ es el factor de degradación asociado al proceso de transmisión del flujo

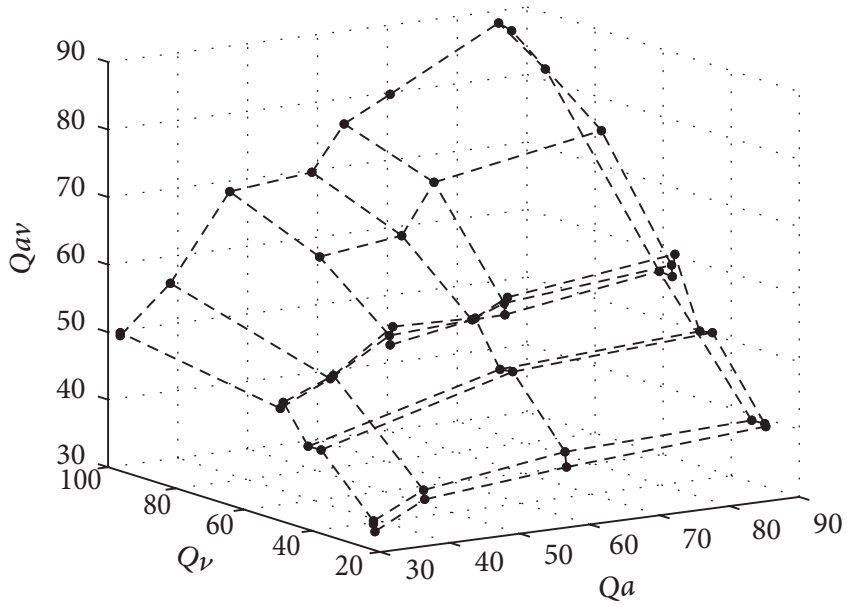


Figura 3.6: Calidad audiovisual en función de la calidad de los flujos de audio y vídeo [Garcia and Raake, 2009]

de vídeo ($x = V$) o audio ($x = A$).

$$Q_{av} = a - b \cdot Icod_A - c \cdot Icod_V - d \cdot Itra_A - e \cdot Itra_V - f \cdot Icod_A \cdot Icod_V \\ -g \cdot Itra_A \cdot Itra_V - h \cdot Icod_A \cdot Itra_V - i \cdot Icod_V \cdot Itra_A \tag{3.23}$$

Tras aplicar un análisis de regresión múltiple sobre los resultados de los experimentos subjetivos, los autores ofrecen una serie de valores para los parámetros de ajuste (tabla 3.4).

Tabla 3.4: Coeficientes del modelo de García, versión 2009

Parámetros								
a	b	c	d	e	f	g	h	i
88,195	0,379	0,588	0,625	0,625	-0,005	-0,007	-0,011	-0,007

Según los autores, con este modelo consiguieron obtener una correlación del 96 % al predecir la calidad audiovisual a partir de los factores de degradación extraídos de los experimentos subjetivos.

En [Garcia et al., 2011] se refina el modelo anterior, contemplando la influencia de la resolución del vídeo, el tipo de las degradaciones y el tipo de contenido. Para ello, utilizando los resultados de los experimentos subjetivos, llevan a cabo análisis de regresión múltiple independientes para los distintos tipos de contenido considerados

(tabla 3.5).

Tabla 3.5: Tipos de contenido contemplados en el modelo de García, versión 2011

ID	Vídeo	Audio
A	Tráiler de película	Conversación sobre música
B	Entrevista	Conversación
C	Partido de fútbol	Conversión sobre ruido
D	Película	Música clásica
E	Vídeo musical	Música pop

Para el caso del modelo basado en componentes de calidad, los resultados del ajuste del modelo se muestran en la tabla 3.6. Como se puede ver en dicha tabla, el vídeo es la componente predominante, especialmente para contenidos High Definition (HD). En el caso de contenidos Standard Definition (SD), las calidades de audio y vídeo están más equilibradas, por lo que tanto β como α son iguales a 0. En cuanto a la dependencia con el contenido, se puede ver por ejemplo, en el caso de la secuencia "HD E", la importancia equilibrada del audio y el vídeo al tratarse de un vídeo musical.

Tabla 3.6: Coeficientes del modelo basado en componentes de calidad de García et al, versión 2011

Secuencia	α	β	γ	μ
HD global	28,49	0	0,13	0,006
HD A	24,57	0	0,28	0,006
HD B	27,50	0	0,11	0,006
HD C	24,37	0	0,21	0,005
HD D	27,85	0	0,17	0,005
HD E	32,59	0	0	0,007
SD global	30,99	0	0	0,006
SD A	32,77	0	0	0,006
SD B	30,21	0	0	0,006
SD C	25,83	0	0,15	0,005
SD D	32,06	0	0	0,006
SD E	30,86	0	0	0,006

Para el caso del modelo basado en degradaciones de calidad, los resultados del ajuste del modelo se presentan en la tabla 3.7.

Una extensión de este modelo, descrita en [Garcia et al., 2013] y estandarizada por ITU en la recomendación ITU-T P.1201.2 [ITU, 2012e], se basa en combinar los dos enfoques: el enfoque basado en factores de calidad y el enfoque basado en factores de degradación, dando un peso de 0,7 a la primera componente y un peso de 0,3 a la segunda, como se puede ver en la ecuación 3.24. Los parámetros de ajuste de este

Tabla 3.7: Coeficientes del modelo basado en factores de degradación de García et al, versión 2011

Secuencia	a	b	c	d	e	f	g	h	i
HD global	94,33	0,466	0,713	0,652	0,712	-0,008	-0,007	-0,007	-0,009
HD A	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
HD B	94,33	0,539	0,814	0,752	0,727	-0,010	-0,009	-0,008	-0,017
HD C	94,33	0	0,786	0,685	0,724	0	-0,007	0	-0,012
HD D	94,33	0,416	0,851	0,601	0,724	-0,007	-0,007	-0,007	-0,013
HD E	94,33	0,560	0,519	0,711	0,667	-0,009	-0,008	-0,009	-0,011
SD global	82,90	0,387	0,511	0,539	0,507	-0,004	-0,005	-0,006	-0,006
SD A	82,90	0,333	0,411	0,471	0,523	0	-0,004	-0,008	0
SD B	82,90	0,510	0,521	0,677	0,522	-0,006	-0,004	-0,007	-0,012
SD C	82,90	0	0,657	0,567	0,462	0	-0,002	0	-0,010
SD D	82,90	0,324	0,472	0,559	0,492	-0,004	-0,005	-0,004	-0,005
SD E	82,90	0,309	0,398	0,613	0,484	0	-0,006	0	-0,007

modelo se presentan en la tabla 3.8.

$$\begin{aligned} Q_{av} = 0,7 \cdot (\alpha + \beta \cdot Q_a + \gamma \cdot Q_v + \mu \cdot Q_a \cdot Q_v) + \\ +0,3 \cdot (a - b \cdot Icod_A - c \cdot Icod_V - d \cdot Itra_A - e \cdot Itra_V - f \cdot Icod_A \cdot Icod_V \\ -g \cdot Itra_A \cdot Itra_V - h \cdot Icod_A \cdot Itra_V - i \cdot Icod_V \cdot Itra_A) \end{aligned} \quad (3.24)$$

Tabla 3.8: Coeficientes del modelo ITU-T P.1201.2

α	β	γ	μ	a	b	c	d	e	f	g	h	i
5,89	0	0,52	0,0045	100	0,32	0,9	0,705	1,02	0	-0,007	-0,008	-0,01

Conclusiones extraídas del estado del arte Como puede extraerse del análisis realizado, las conclusiones de los distintos estudios son heterogéneas, debido principalmente a las diferencias entre aplicaciones y las condiciones en las que se llevaron a cabo los experimentos. Sin embargo, se pueden extraer una serie de conclusiones comunes a todos estos estudios:

- La calidad percibida de una secuencia audiovisual está determinada principalmente por la calidad del vídeo.
- La interacción entre el audio y el vídeo no es despreciable.
- La interacción entre el audio y el vídeo y su influencia sobre la calidad audiovisual dependen del tipo de aplicación y del tipo de contenido de la secuencia.
    - Cuanto más compleja es una componente, mayor es su peso.

La mayoría de los modelos presentados son casos particulares de un modelo general (ecuación 3.25).

$$Q_{av} = \alpha + \beta \cdot Q_a + \gamma \cdot Q_v + \mu \cdot Q_a \cdot Q_v \tag{3.25}$$

3.4.1.2. Modelo propuesto

Tras analizar las distintas propuestas disponibles en la literatura de esta área, se ha tomado la decisión de tomar como base el modelo propuesto en [Garcia et al., 2013] y estandarizado por ITU en la recomendación ITU-T P.1201.2 [ITU, 2012e] y adaptarlo a las particularidades del contexto considerado en esta tesis.

La expresión del modelo viene dada por la ecuación 3.24 con los parámetros de ajuste de la tabla 3.8.

Aunque en su concepción el modelo está orientado a ser un modelo híbrido, que tiene en cuenta tanto valoraciones de calidad como factores de degradación, lo cierto es que el modelo se puede escribir por completo en términos de factores de degradación, ya que las valoraciones de calidad están definidas según la ecuación 3.26.

$$Q_x = 100 - Icod_x - Itra_x \tag{3.26}$$

Por otro lado, para el caso particular de esta tesis, se considera $Itra_x = 0$, por lo que el modelo se reduce a la ecuación 3.27, con los parámetros de ajuste de la tabla 3.9.

$$Q_{av} = 0,7 \cdot (\alpha + \gamma \cdot Q_v + \mu \cdot Q_a \cdot Q_v) + 0,3 \cdot (a - b \cdot Icod_a - c \cdot Icod_v) \tag{3.27}$$

Tabla 3.9: Coeficientes del modelo de calidad audiovisual propuesto (adaptación de ITU-T P.1201.2)

α	γ	μ	a	b	c
5,89	0,52	0,0045	100	0,32	0,9

Así pues, para poder aplicar este modelo, únicamente es necesario disponer de los modelos adecuados con los que expresar $Icod_v$ e $Icod_a$.

Calidad de vídeo La propia recomendación ITU-T P.1201.2 ofrece modelos para estimar el efecto de la codificación tanto para vídeo como para audio. El modelo de vídeo se presenta en la ecuación 3.28, con los parámetros de ajuste de la tabla 3.10.

$$Icod_v = a_{1v} \cdot e^{a_{2v} \cdot BitPerPixel} + a_{3v} \cdot ContentComplexity + a_{4v} \tag{3.28}$$

Tabla 3.10: Parámetros de ajuste del modelo de vídeo ITU-T P.1201.2

Resolución	**Parámetros**			
	a_{1v}	a_{2v}	a_{3v}	a_{4v}
SD	61,28	-11	6	6,21
HD	51,28	-22	6	6,21

La estimación de la complejidad del contenido se realiza mediante la ecuación 3.29.

$$ContentComplexity = \frac{\sum_{SC} Nw}{\sum_{SC} s^I_{sc} \cdot Nw} \cdot \frac{PixelPerFrame \cdot FrameRate}{1000} \tag{3.29}$$

En esta ecuación, s^I_{sc} es un vector que contiene el tamaño medio por escena de las tramas I, es decir $s^I_{sc} = (s^I_{sc1}, s^I_{sc2}, s^I_{sc3}, ...) \in IR^S$, donde S es el número de escenas en la ventana de medida y s^I_{sci} es el tamaño medio de las tramas I en la escena i (ignorando la primera trama I).

Nw se calcula de la siguiente manera: si N es un vector S-dimensional que contiene el número de GoP por escena, es decir, $N = (n_{sc1}, n_{sc2}, n_{sc3}, ...) \in IR^S$, y si m es el índice de la escena con menor valor de s^I_{sci}, y s es el índice de la escena, entonces:

$$Nw(s) = \begin{cases} N(s) \cdot 16, & s = m \\ N(s), & s \neq m \end{cases} \tag{3.30}$$

Una de las premisas del modelo ITU-T P.1201.2 es que los parámetros de entrada se puedan extraer de la información contenida en las cabeceras de los paquetes de los flujos de transporte, tanto para flujos no cifrados como para flujos cifrados. Dependiendo del nivel de cifrado algunos parámetros tendrán que ser estimados de acuerdo a lo recogido en la recomendación, ya que no podrán leerse directamente del flujo de bits.

Como se puede ver, el modelo no analiza directamente el contenido (trama a trama y pixel a pixel) de las tramas de vídeo para obtener una estimación del nivel de complejidad del contenido, reduciéndose solo a parámetros de codificación como el tamaño medio de las tramas I, el tamaño del GoP, etc.

Así pues, **esta limitación en el modelo sirve como motivación para el estudio más profundo de un modelo de estimación de calidad de vídeo que contemple el efecto de la codificación y lleve a cabo un análisis más exhaustivo de la complejidad espacial y temporal del contenido**. El desarrollo del modelo de calidad de vídeo se lleva a cabo en el capítulo 4.

Calidad de audio En cuanto a la degradación en el audio, el modelo que se utilizará en esta tesis es el recomendado por ITU-T P.1201.2, el cual se recoge en la ecuación

3.31.

$$Icod_a = a_{1a} \cdot e^{a_{2a} \cdot BitRate} + a_{3a} \tag{3.31}$$

Los parámetros de ajuste del modelo de audio se presentan en la tabla 3.11.

Tabla 3.11: Parámetros de ajuste del modelo de audio ITU-T P.1201.2

Códec	**Parámetros**		
	a_{1a}	a_{2a}	a_{3a}
MP2	100	-0,02	15,48
AC-3 (Dolby Digital)	100	-0,03	15,70
AAC-LC	100	-0,05	14,60
HE-AAC	100	-0,11	20,06

3.4.2. Sincronización audio-vídeo

De manera análoga a la propuesta de [de la Cruz Ramos, 2012], para la estimación del factor de degradación asociado a la desincronización entre los flujos de audio y vídeo (I_{ls}) se podría aplicar una variación logarítmica entre los umbrales de detección y aceptabilidad especificados en ITU-R BT.1359-1 [ITU, 1998b], basada en la hipótesis de que la calidad decae rápidamente una vez que el retardo o adelanto entre flujos es detectable. Así pues, I_{ls} se podría estimar mediante la ecuación 3.32.

$$I_{ls} = \begin{cases} 100, & T \leq -90ms \\ 332,25 \cdot log(-T) - 549,25, & -90ms < T < -45ms \\ 0, & -45ms \leq T \leq 125ms \\ 587,25 \cdot log(T) - 1231,5, & 125ms < T < 185ms \\ 100, & T \geq 185ms \end{cases} \tag{3.32}$$

Como se puede ver en la figura 3.7, si la desincronización entre los flujos de audio y vídeo supera los umbrales de aceptabilidad, la degradación alcanza su valor máximo, mientras que si la desincronización es menor que los umbrales de detección, la degradación es nula.

Sin embargo, además de ITU, otros organismos de estandarización han propuesto umbrales de detección y aceptabilidad distintos. Por ejemplo, la recomendación R37 de European Broadcasting Union (EBU) establece unos umbrales de -40 ms y +60 ms para programas de televisión [EBU, 2007]. Por su parte, Advanced Television System Committee (ATSC) argumenta en ATSC IS-191 [ATSC, 2003] que la diferencia temporal entre los flujos de audio y vídeo no debería exceder nunca de -15 ms y +40 ms. Estos umbrales están respaldados por DSL Forum [DSL, 2006] y por ITU-T G.1080

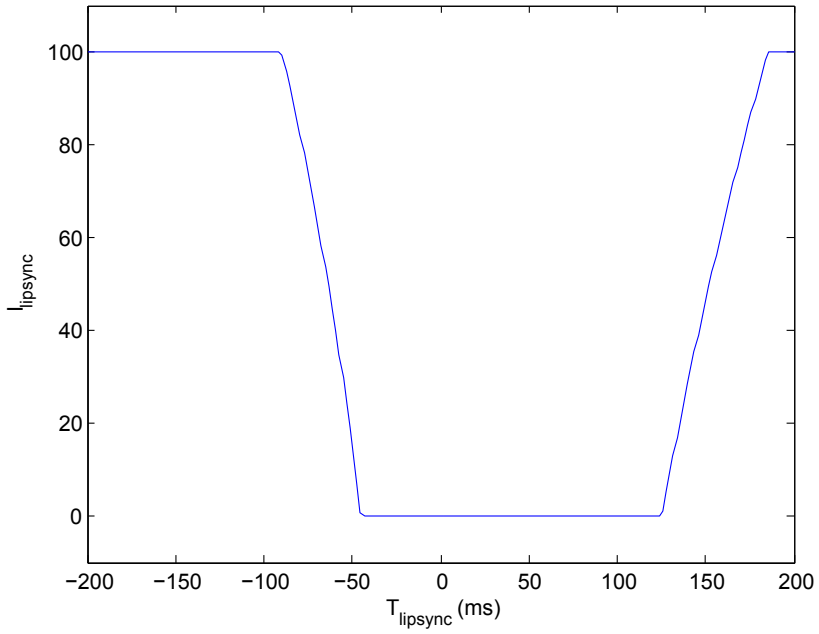


Figura 3.7: Factor de degradación de calidad asociado al lipsync

[ITU, 2008b].

A la vista de las distintas opiniones en cuanto a los umbrales de detección y aceptabilidad, parece razonable plantear la hipótesis de que la degradación de la desincronización entre los flujos de audio y vídeo tendrá más o menos efecto dependiendo del tipo de contenido de la secuencia de vídeo. Para modelar este efecto se podría actuar de dos maneras distintas:

- Ponderar el efecto del contenido añadiendo un factor multiplicativo a I_{ls}
- Incluir el efecto del contenido en la estimación de I_{ls}

En esta tesis se ha optado por incluir el efecto del contenido en la propia estimación de I_{ls}. Para ello, se reescribe la ecuación 3.32 de forma paramétrica (ecuación 3.33).

$$I_{ls} = \begin{cases} 100, & T \leq A_1 \\ \alpha \cdot log(-T) + \beta, & A_1 < T < D_1 \\ 0, & D_1 \leq T \leq D_2 \\ \gamma \cdot log(T) + \xi, & D_2 < T < A_2 \\ 100, & T \geq A_2 \end{cases} \tag{3.33}$$

En la ecuación 3.33 A_i y D_i son los umbrales de aceptabilidad y detección que variarán en función del contenido. Para el ajuste de estos coeficientes, el enfoque propuesto se basa en realizar una clasificación de distintos tipos de contenido donde el efecto del lipsync es similar (tabla 3.12).

Tabla 3.12: Efecto de la sincronización audio-vídeo en función del contenido

Categoría	**Tipos de secuencia**	**Efecto del lipsync**
Programas tipo busto parlante	Noticiarios, programas de opinión, entrevistas, etc.	Muy alto
Programas mixtos (diálogos, sonidos en off)	Películas, videos musicales, etc.	Alto
Programas con voz en off	Retransmisiones de eventos deportivos, documentales, etc.	Moderado

Como aproximación, se proponen los valores de la tabla 3.13 para los umbrales de aceptabilidad y detección en función del tipo de contenido:

Tabla 3.13: Umbrales aproximados de aceptabilidad y detección del lipsync en función del tipo de contenido

Categoría	**Umbrales (ms)**			
	A_1	D_1	D_2	A_2
Programas tipo busto parlante	-80	-40	115	165
Programas mixtos (diálogos, sonidos en off)	-90	-45	125	185
Programas con voz en off	-100	-55	140	200

Una vez conocidos A_i y D_i, el resto de coeficientes del modelo se pueden calcular con las siguientes ecuaciones.

$$\alpha = \frac{100}{log\left(\frac{A_1}{D_1}\right)} \tag{3.34}$$

$$\beta = -\alpha \cdot log(-D_1) \tag{3.35}$$

$$\gamma = \frac{100}{log\left(\frac{A_2}{D_2}\right)} \tag{3.36}$$

$$\xi = -\gamma \cdot log(-D_2) \tag{3.37}$$

3.4.3. Degradación de calidad debida a la transmisión

Como se comentó al hablar del efecto de los componentes continuos, la red y los protocolos utilizados para la transmisión del vídeo mediante Internet, conlleva una serie de degradaciones que afectan directamente a la calidad percibida por los usuarios.

Las degradaciones que se han tenido en cuenta en el desarrollo del modelo son las siguientes: tiempo de buffering inicial, tiempo total de rebuffering y número de eventos de rebuffering. Además, se contempla el efecto que tienen los cambios en el nivel de calidad que se puede producir como respuesta por parte de los algoritmos de adaptación a las condiciones cambiantes de la red.

Debido a la importancia y a la extensión necesaria para desarrollar el modelo de I_{tra}, éste se presenta en el capítulo 5 de manera independiente.

3.5. Componentes puntuales

3.5.1. Cambio de canal

Es evidente que el tiempo de cambio de canal es un factor que influye en la valoración que los usuarios hacen de un servicio. Por esta razón, en la literatura científica se pueden encontrar múltiples artículos que tratan sobre el tiempo de cambio de canal. Son muy comunes por ejemplo los trabajos orientados a la reducción del tiempo de cambio de canal en servicios como IPTV. Sin embargo, no son tan abundantes los trabajos orientados a cuantificar o a estimar el efecto en la calidad percibida del tiempo de cambio de canal.

3.5.1.1. Revisión del estado del arte

Aunque la mayoría de trabajos relacionados con el tiempo de cambio de canal de los últimos años se han realizado en el contexto de sistemas IPTV, los conceptos subyacentes de los mismos pueden ser aplicados a sistemas OTT, por lo que se considera adecuado tenerlos en cuenta en esta tesis.

Optimización del tiempo de cambio de canal En [Asghar et al., 2009] se lleva a cabo un trabajo orientado a mejorar la calidad de experiencia en IPTV mediante la mejora de distintos aspectos del servicio. Uno de los aspectos considerados es el tiempo de cambio de canal. Los autores consideran que los tiempos de cambio de canal en IPTV son mayores que los correspondientes en sistemas convencionales debido a los siguientes factores:

- Retardos en la señalización Internet Group Management Protocol (IGMP)
- Tiempo de decodificación The Moving Picture Experts Group (MPEG)
- Tiempo de adquisición de la primera trama clave
- Tiempo de adquisición de claves del sistema de acceso condicional

En [Banodkar et al., 2008] se propone un mecanismo alternativo al cambio de canal instantáneo tradicional de IPTV (que se basa en reducir la latencia del tiempo de cambio de canal utilizando un canal unicast) basado en multicast. Más concretamente se propone utilizar un flujo multicast secundario de menor calidad, en el que el tiempo unión es menor, mientras en paralelo se produce la verdadera unión multicast al flujo de alta calidad.

En [Siebert et al., 2009] se lleva a cabo una revisión de las últimas técnicas aplicadas en IPTV para mejorar el tiempo de cambio de canal, entre las que destacan las siguientes:

- Reducción del GoP
- Reducción del tiempo de buffering inicial del vídeo
- Utilización de un canal auxiliar con menor GoP para el cambio de canal
- Utilización de réplicas (sub-canales) para minimizar el tiempo de espera para obtener una trama I
- Utilización de flujos unicast para el cambio de canal
- Cambio rápido de canal utilizando codificación escalable
- Cambio rápido de canal utilizando tramas SI/SP en H.264/Advanced Video Coding (AVC)

En 2011, el IETF en la RFC 6285 [IETF, 2011] estandarizó un mecanismo basado en la utilización de un canal unicast para realizar un cambio de canal rápido en sesiones multicast basadas en RTP.

En [Ramos et al., 2011] se presenta un enfoque predictivo para abordar el problema del cambio de canal. Los autores afirman que la mayoría de los usuarios realizan los cambios de canal de manera lineal, navegando hacia arriba o hacia abajo en la lista de canales. Teniendo en cuenta este comportamiento, los autores proponen que durante los periodos de zapping, los Set-Top Box (STB) de los usuarios se vaya uniendo a los canales vecinos, con el objetivo de minimizar el tiempo de cambio de canal cuando los usuarios realicen múltiples cambios de canal en un periodo de tiempo limitado.

En [Van Wallendael et al., 2012] se propone utilizar una configuración de codificación de vídeo escalable o Scalable Video Coding (SVC) que hace posible, mediante una capa básica y una capa de refinado, mejorar el tiempo de cambio de canal, sin afectar en el ancho de banda utilizado.

Efecto del cambio de canal en la QoE En [Kooij et al., 2006] se lleva a cabo un estudio enfocado a analizar el efecto que tiene el tiempo de cambio de canal en la calidad percibida. Para ello, los autores se basan en las ideas presentadas en ITU-T G.1030 [ITU, 2014a], donde se estudia la calidad percibida en la navegación web en función de los tiempos de respuesta y de descarga. De manera análoga al modelo expuesto en dicha recomendación, los autores proponen una variación logarítmica entre dos valores extremos de tiempo de cambio de canal (T_z), como se puede ver en la ecuación 3.38.

$$MOS_z = MOS_{max} + (MOS_{max} - MOS_{min}) \cdot \frac{ln(T_z) - ln(T_{min})}{ln(T_{min}) - ln(T_{max})} \tag{3.38}$$

Para obtener dichos valores extremos, los autores se basan en las ideas de [Nielsen, 1994]:

- 0,1 segundos es el límite para considerar el cambio de canal como instantáneo.
- 1 segundo es el límite para no interrumpir el "flujo de pensamiento", aunque no haya sensación de reacción instantánea por parte del sistema.
- 10 segundos es el límite para mantener la atención del usuario.

Así pues, los autores seleccionaron los siguientes parámetros para el modelo: $MOS_{max} = 5$, $MOS_{min} = 1$, $T_{max} = 5s$, $T_{min} = 0,1s$.

Sustituyendo en la ecuación 3.38 se obtiene:

$$MOS_z = max\{min\{-1,02 \cdot ln(T_z) + 2,65; 5\}; 1\} \tag{3.39}$$

Como se puede extraer de la ecuación 3.39, los valores de tiempo de cambio de canal necesarios para garantizar una MOS de al menos 3,5, deben ser menores de 0,43 segundos. La validación de este modelo se hizo mediante tests subjetivos, obteniendo una correlación de 0,99 entre la predicción del modelo y los resultados de los tests.

En [Kooij et al., 2009b] los autores del modelo anterior llevaron a cabo una revisión de dicho modelo, considerando tests subjetivos más acordes a los escenarios típicos de consumo de televisión: cambio de canal mediante mando a distancia y posición de visionado relajada. Los tests llevados a cabo en [Kooij et al., 2006] fueron realizados mediantes ordenadores personales y el cambio de canal se simulaba mediante botones en una página web. Además, los autores estudiaron el efecto de cambiar de canal pulsando el número concreto del nuevo canal, o por el contrario, utilizando los botones de subir/bajar canal, no encontrando grandes diferencias entre ambas opciones.

Así pues, estos nuevos tests dieron como resultado un modelo algo menos exigente, como se puede ver en la siguiente ecuación 3.40.

$$MOS_{z,var=0} = \begin{cases} -2,1 \cdot T_z + 4,92, & 0 \leq T_z \leq 1,04 \\ -1,11 \cdot ln(T_z) + 2,78, & 1,04 \leq T_z \leq 4,97 \\ 1, & 4,97 \leq T_z \end{cases} \tag{3.40}$$

Como se puede ver en la figura 3.8, con este nuevo modelo el tiempo necesario para obtener una MOS de 3,5 se relaja un poco, siendo ahora de 0,67 segundos.

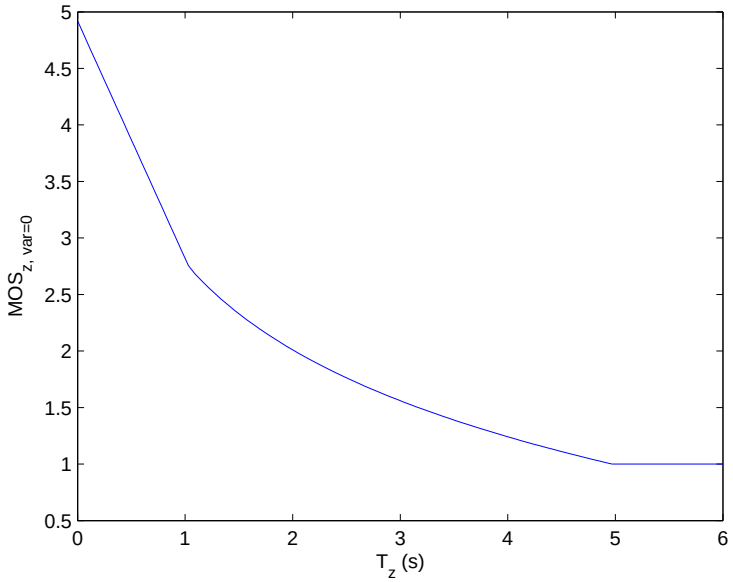


Figura 3.8: Calidad del cambio de canal con varianza nula

Además, los autores descubrieron que la varianza de los tiempos de cambio de canal afecta negativamente a la percepción de la calidad, por lo que incluyen en su modelo un factor de corrección de acuerdo a la ecuación 3.41.

$$\Delta MOS_z = \begin{cases} var(T_z), & E[T_z] < 0,42 \\ \frac{var(T_z)}{E[T_z]}, & E[T_z] \geq 0,42 \end{cases} \tag{3.41}$$

Así pues, para aplicar este modelo, en primer lugar se utiliza la fórmula general, que asume varianza 0 y después se aplica el decremento de MOS en función del valor de la varianza. Por último, para asegurar que el resultado está dentro de los márgenes aceptables de MOS, se aplica una expresión de recorte, como se puede ver en la ecuación

3.42.

$$MOS_z = max\{MOS_{z,var=0} - \Delta MOS_z; 1\} \tag{3.42}$$

Existen otros trabajos, dentro del marco de la calidad percibida, orientados a mejorar la percepción del cambio de canal, introduciendo contenido auxiliar mientras se produce dicho cambio.

Por ejemplo, en [Kooij et al., 2009a] y en [Godana et al., 2009] se afirma que si, mientras se produce el cambio de canal, se introducen anuncios o pequeños clips de vídeo o información del contenido que se va a ver, se puede mejorar la QoE. Según los autores, esta técnica funciona cuando el tiempo de cambio de canal es largo, ya que los usuarios prefieren poder ver algún contenido frente a ver una "pantalla negra". Teniendo esto en cuenta, los autores proponen un sistema que en función de una estimación del tiempo de cambio de canal, muestra una pantalla negra, un pequeño vídeo o una foto.

Un experimento similar se llevó a cabo en [Kooij and Geijer, 2012], utilizando un juego para amenizar la espera del cambio de canal. Según los autores, si los tiempos de espera superan los 2,25 segundos, introducir un juego mejora la QoE. Más concretamente, en escenarios con tiempos de espera de 3 segundos, utilizando un juego se consigue una MOS mayor que 3,5. Sin embargo, si el tiempo de espera es menor de un segundo, introducir un juego no mejora, sino que empeora la calidad de experiencia (debido a que no hay tiempo para conseguir jugar, lo cual en general frustra al usuario).

3.5.1.2. Modelo propuesto

El modelo seleccionado para utilizar en esta tesis se basa en el presentado en [Kooij et al., 2009b]. Como se ha comentado anteriormente, este modelo simula unas condiciones de visionado típicas de servicios de televisión, donde el usuario está cómodamente sentado y realiza el cambio de canal utilizando un mando a distancia. Además, tiene en cuenta que la varianza en los tiempos de cambio de canal afecta negativamente a la calidad de experiencia.

Sin embargo, como se comentó en anteriormente, en el modelo global de esta tesis, el efecto del tiempo de cambio de canal se modela como una degradación por lo que el valor estimado de MOS_z debe ser convertido a un valor de degradación I_z.

Por otro lado, los valores extremos de MOS_z que proporciona el modelo deben ser adaptados a los valores máximo y mínimos de MOS considerados en [ITU, 2012e].

Por tanto, la ecuación 3.40 quedaría de la siguiente manera:

$$MOS_{z,var=0} = \begin{cases} -2,1 \cdot T_z + 4,9, & 0 \leq T_z \leq 1,04 \\ -1,067 \cdot ln(T_z) + 2,757, & 1,04 \leq T_z \leq 4,97 \\ 1,05, & 4,97 \leq T_z \end{cases} \quad (3.43)$$

En la ecuación 3.42 también se debe modificar el valor de MOS mínimo, resultando:

$$MOS_z = max\{MOS_{z,var=0} - \Delta MOS_z; 1,05\} \quad (3.44)$$

Así pues, el modelo se debe aplicar siguiendo los pasos que se describen a continuación:

1. Estimación del factor de calidad del cambio de canal suponiendo varianza 0 ($MOS_{z,var=0}$), según la ecuación 3.43.
2. Estimación del factor de penalización en función de la varianza (ΔMOS_z), según la ecuación 3.41.
3. Aplicación de una función de recorte para obtener MOS_z, según la ecuación 3.44.
4. Conversión de MOS_z a Q_z (en escala R), mediante la ecuación 3.6.
5. Calcular $I_z = 100 - Q_z$.

En la figura 3.9 se muestra la curva que relaciona el tiempo de cambio de canal T_z con el factor de degradación asociado al mismo I_z, asumiendo varianza nula en el tiempo de cambio de canal.

3.5.2. Acceso aleatorio

En servicios de vídeo bajo demanda, es habitual que el usuario tenga la posibilidad de seleccionar un instante de tiempo al que desea desplazarse para continuar la reproducción desde ahí. Por ejemplo, el usuario podría decidir volver a un instante pasado o avanzar en el tiempo para encontrar alguna escena interesante. Esta característica se suele conocer como "seeking", o "acceso aleatorio". En esta tesis se utilizará el término "acceso aleatorio" para referir dicha funcionalidad.

Debido a que es una característica importante del servicio, en esta tesis se considera su efecto en la calidad de la experiencia global del servicio.

3.5.2.1. Revisión del estado del arte

En la revisión del estado del arte que se ha llevado a cabo no se han encontrado trabajos que estudien el efecto que tiene el tiempo de acceso aleatorio en la calidad

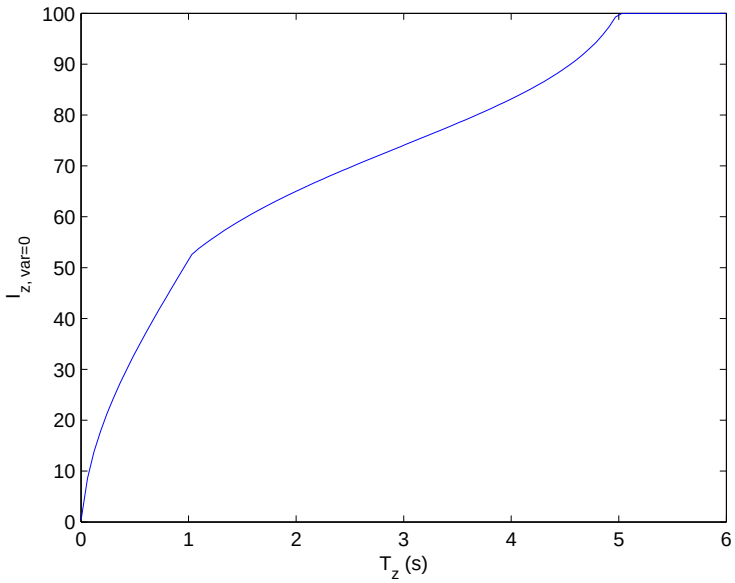


Figura 3.9: Degradación asociada al tiempo de cambio de canal con varianza nula

percibida por el usuario. Hay diversos trabajos que tratan sobre el acceso aleatorio en vídeo, pero con otros enfoques.

Por ejemplo, en [Yang et al., 2009] y [Xu et al., 2010] se proponen mecanismos para implementar el acceso aleatorio en vídeo para sistemas Peer-to-peer (P2P).

Otros artículos se centran en la efectividad de distintas técnicas para buscar información (automáticamente) dentro de un vídeo, lo que se conoce como "vídeo browsing", como por ejemplo [Duan et al., 2004].

En [Tse et al., 1999], [Li et al., 2000], [Hurst et al., 2004] se analizan distintas interfaces de usuario y controles de reproducción para realizar acceso aleatorio y otras operaciones básicas (pausa, avance rápido, etc.).

En ITU-T G.1080 [ITU, 2008b] se destaca la importancia de una baja latencia en los modos de reproducción (stop, pausa, avance rápido, etc.) en servicios IPTV. Sin embargo, no proporciona valores concretos para dicha latencia: "as each trick feature latency directly affects QoE, the latency is required to be sufficiently low to meet user's requirement for QoE related to VoD trick features".

3.5.2.2. Modelo propuesto

Debido a la falta de literatura, se ha decidido desarrollar un modelo propio para evaluar el efecto que tiene la funcionalidad de acceso aleatorio en la calidad percibida por el usuario.

Para desarrollar este modelo, es interesante conocer en primer lugar los factores que influyen en el tiempo y en la precisión de los accesos aleatorios en flujos de vídeo. Este análisis se va a hacer desde dos puntos de vista, desde el punto de vista del sistema de transmisión (MPEG-DASH) y desde el punto de vista de la codificación (centrando el análisis en H.264).

MPEG-DASH El estándar MPEG-DASH ofrece algunas recomendaciones para implementar la funcionalidad de seeking o acceso aleatorio. Mediante los ficheros MPD, el cliente tiene acceso al instante de tiempo en el que comienza cada segmento, por lo que al realizar un acceso aleatorio al instante T_M, el segmento que con más probabilidad contendrá las muestras del contenido asociado al instante T_M será el segmento $S[i]$, siendo i el máximo valor que cumple $S[i].StartTime \leq T_M$.

Sin embargo, se debe tener en cuenta que la información temporal recogida en el MPD puede ser aproximada, debido a una serie de factores: posición de los Stream Access Point (SAP), alineado de las pistas de medios y derivas en la temporización de las pistas. Debido a esta falta de precisión, puede que el segmento $S[i]$, identificado como candidato a contener el instante T_M, comience después de dicho instante, siendo el segmento correcto $S[i-1]$. En este caso, el estándar contempla dos opciones: actualizar el instante de reproducción al instante de tiempo que contiene la primera muestra del segmento $S[i]$ o bien solicitar el segmento $S[i-1]$. Si se elige la primera opción, habrá un pequeño error en el instante de tiempo desde el que se reanuda la reproducción. Si se elige la segunda opción, no habrá error en el instante desde el que se reanuda la reproducción, pero el tiempo de acceso aleatorio aumentará, debido a que se tiene que solicitar un segmento extra para corregir dicho error.

Un concepto importante son los denominados SAP o Puntos de Acceso Aleatorio. Un SAP se define como una posición en una representación que permite comenzar la reproducción de un flujo de medios usando solo información contenida en dicha representación a partir de dicha posición y opcionalmente datos de inicialización [ISO, 2014a].

Además de seleccionar el segmento adecuado, para llevar a cabo un acceso aleatorio preciso al instante T_M el cliente MPEG-DASH necesita acceder a un SAP. Para ello, el cliente puede consultar un "Segment Index" u otras señales que se pueden incluir en el fichero MPD para obtener información adicional que le ayude a localizar los SAP dentro de un segmento. Dependiendo de donde se encuentre el SAP el cliente de nuevo tiene dos opciones similares a las anteriores: empezar a decodificar y renderizar desde el SAP anterior más próximo a T_M, con lo cual, se asume un cierto error en el instante de tiempo desde el que se reanuda la reproducción; o bien, empezar a decodificar desde el SAP anterior más próximo a T_M y no reanudar el renderizado hasta alcanzar T_M, con lo cual, se asume un cierto retardo en la reanudación de la reproducción.

H.264 Por su parte, el estándar H.264 ofrece varias funcionalidades que permite realizar acceso aleatorio a diferentes instantes del flujo de vídeo. A continuación se resume cada una de estas posibilidades:

- Tramas/slices I: son tramas o slices que no necesitan referencia a otras tramas o slices para ser decodificadas. En la codificación de este tipo de tramas se explota la correlación espacial de los pixel de la trama. Las tramas I se utilizan como base en la codificación y decodificación de otras tramas y proporcionan puntos de acceso aleatorio donde se puede llevar a cabo el acceso aleatorio. El número de tramas entre tramas I consecutivas suele marcar el tamaño del GoP, aunque pueden haber varias tramas I en un GoP.

- Tramas/slices SP/SI [Karczewicz and Kurceren, 2003], [Setton and Girod, 2005]: este tipo de tramas y slices se han introducido en el perfil extendido de H.264 y permiten una conmutación eficiente entre flujos de vídeo, además de acceso aleatorio, por lo que su utilidad en mecanismos de streaming adaptativo es indudable. De manera similar a las tramas P, las tramas SP utilizan codificación predictiva mediante compensación de movimiento. La diferencia entre las tramas SP y las tramas P es que las tramas SP permiten la reconstrucción de tramas idénticas, aunque se usen diferentes tramas de referencia. Debido a esta propiedad, las tramas SP se pueden utilizar como una alternativa a las tramas I en diversas aplicaciones como pueden ser: conmutación entre flujos a distintas tasas de bit, acceso aleatorio, fast/back forward y protección de errores. Además, como las tramas SP utilizan compensación de movimiento, suponen un gran ahorro, en cuanto a tasa de bit de codificación, con respecto a las tramas I. Por su parte, las tramas SI se usan de manera similar a las tramas SP, con la salvedad de que las predicciones se realizan en el dominio del espacio (como las tramas I). Este tipo de tramas se pueden utilizar para conmutar de una secuencia a una secuencia completamente distinta (donde no es beneficioso usar compensación de movimiento), por lo que son de especial interés para llevar a cabo operaciones de acceso aleatorio y corrección de errores.

- Unidades de acceso Instantaneous Decoding Refresh (IDR) [Wiegand et al., 2003]: una de las estructuras que define la capa de abstracción de red o Network Abstraction Layer (NAL) se denominan "secuencias de vídeo codificado" (Coded Video Sequences). Una secuencia de vídeo codificado consiste en un conjunto de unidades de acceso (conjunto de unidades NAL cuya decodificación resulta en una imagen decodificada) secuenciales dentro del flujo de unidades NAL y que utiliza un único conjunto de parámetros de secuencia. Cada una de estas secuencias de vídeo codificado pueden ser decodificadas de manera independiente, dada la in-

formación del conjunto de parámetros necesario. Al inicio de cada secuencia de vídeo codificado se incluye una unidad de acceso de refresco de decodificación instantánea IDR. Una unidad de acceso IDR contiene una imagen intra (una imagen codificada que puede ser decodificada sin decodificar ninguna imagen en el flujo de unidades NAL). Además, la presencia de una unidad de acceso IDR indica que ninguna imagen posterior a dicha unidad de acceso IDR necesitará referencia anterior a la imagen intra que contiene. Expresado de manera más simple, una IDR es un tipo especial de trama I en H.264 que especifica que ninguna trama después de la IDR puede referenciar a tramas anteriores a la IDR

Definición del modelo Teniendo en cuenta estas dos componentes, y de manera análoga al proceso seguido para el cambio de canal, se propone la ecuación 3.45, donde MOS_{aa} es la valoración de calidad de la funcionalidad de acceso aleatorio, $MOS_{aa,error=0}$ es la valoración de calidad de la funcionalidad de acceso aleatorio teniendo en cuenta el tiempo necesario para llevarla a cabo y ΔMOS_{aa} es un factor de degradación asociado al error entre el instante de tiempo objetivo y el instante de tiempo de reinicio de la reproducción.

$$MOS_{aa} = max\{MOS_{aa,error=0} - \Delta MOS_{aa}; 1,05\} \tag{3.45}$$

Para el usuario, desde el punto de vista de la calidad percibida, sería razonable suponer que el tiempo de acceso aleatorio tenga asociado una MOS similar a la del tiempo de cambio de canal. Por tanto, la expresión para la componente de calidad asociada al tiempo de acceso aleatorio se puede definir de manera análoga a [Kooij et al., 2009b], siendo T_{aa} el tiempo necesario para llevar a cabo el acceso aleatorio:

$$MOS_{aa,error=0} = \begin{cases} -2,1 \cdot T_{aa} + 4,9, & 0 \leq T_{aa} \leq 1,04 \\ -1,067 \cdot ln(T_{aa}) + 2,757, & 1,04 \leq T_{aa} \leq 4,97 \\ 1,05, & 4,97 \leq T_{aa} \end{cases} \tag{3.46}$$

Para el factor de degradación asociado a la precisión del acceso aleatorio se propone la utilización de una expresión de la forma $\Delta MOS_{aa} = f(|T' - T|)$, donde $|T' - T|$ es la diferencia temporal entre el instante al que se deseaba acceder y el instante en el que el reproductor inició la reproducción.

Un ejemplo de función f podría ser la proporcionada en la ecuación 3.47 y en la figura 3.10. En cualquier caso, sin contar con datos experimentales no se puede predecir correctamente el aspecto de dicha función, por lo que esta tarea se propone como una

línea de trabajo futuro.

$$\Delta MOS_{aa} = 1 + \frac{4}{1 + e^{-\frac{1}{2}\cdot(|T'-T|-15)}} \tag{3.47}$$

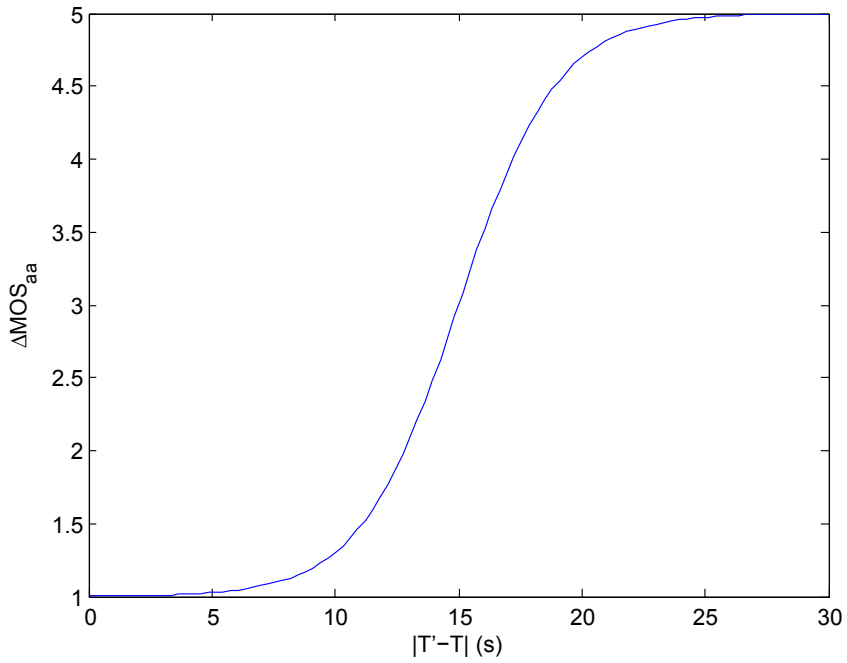


Figura 3.10: Degradación de calidad asociada al error en el acceso aleatorio

Finalmente, habría que transformar el valor de MOS_{aa} en un factor de degradación I_{aa} de manera análoga al caso del cambio de canal:

1. Conversión de MOS_{aa} a Q_{aa} (en escala R), mediante la ecuación 3.6.

2. Calcular $I_{aa} = 100 - Q_{aa}$.

3.6. Resumen y conclusiones

En este capítulo se ha desarrollado un modelo general de estimación de la calidad percibida por los usuarios en un servicio de streaming multimedia OTT.

Este modelo combina las aportaciones de los distintos componentes del servicio, distinguiendo entre componentes continuos y componentes puntuales en función de la fracción de tiempo de prestación del servicio en la que cada componente está presente. Además, dependiendo de la naturaleza de cada componente de servicio, su efecto en la calidad global se modela como un factor de calidad o como un factor de degradación.

En resumen, el modelo propuesto se presenta en la ecuación 3.48.

$$Q = Q_{av} - I_{ls} - I_{tra} - p_z \cdot f(Q_C) \cdot I_z - p_{aa} \cdot f(Q_C) \cdot I_{aa} \tag{3.48}$$

Aunque ya se han introducido a lo largo del capítulo, hay dos factores del modelo que se desarrollarán con mayor detalle en los siguientes capítulos, debido a su importancia. En concreto, en el capítulo 4 se desarrolla un nuevo modelo de estimación de calidad de vídeo, mientras que en el capítulo 5 se desarrolla el modelo que estima la degradación de la calidad introducida por la red (I_{tra}).

Capítulo 4

Modelo de estimación de calidad de vídeo

4.1. Introducción y motivación

En este capítulo se presenta el modelo de estimación de calidad de vídeo que se ha desarrollado en el ámbito de esta tesis, como respuesta a las necesidades concretas del servicio considerado en la tesis y que no se han podido cubrir con los modelos actuales de la literatura.

En primer lugar, no se han encontrado modelos sin referencia No Reference (NR) estandarizados para resoluciones HD, ni tampoco se han encontrado modelos NR que emulen a algún modelo de referencia completo Full Reference (FR) estandarizado. Por otro lado, como se podrá ver en el estudio del estado del arte (sección 4.2), el resto de los modelos analizados no son directamente aplicables a esta tesis por los siguientes motivos:

- La mayor parte de los modelos de calidad de video de la literatura están entrenados utilizando secuencias de vídeo de baja resolución, debido principalmente a que en el momento de ser publicados, las resoluciones utilizadas habitualmente en la mayoría de servicios de vídeo no eran tan altas como las actuales.

- La utilización de pocas secuencias de vídeo de entrenamiento en el desarrollo de modelos de calidad de vídeo hace que la validez y la aplicabilidad de algunos modelos sea limitada. Utilizando un conjunto más numeroso de secuencias de vídeo se puede desarrollar un modelo de calidad más robusto, entrenado utilizando un mayor conjunto de tipos de contenido y degradaciones.

Teniendo en cuenta la motivación descrita, el objetivo de este capítulo es: **desarrollar un modelo sin referencia de calidad percibida para contenidos de vídeo.**

Más concretamente, este modelo se centra en el siguiente escenario:

- **Resolución: Vídeo Full-HD 1920x1080**. La resolución Full-HD es el objetivo al que los distintos proveedores de vídeo OTT irán apuntando en el futuro cercano. Aunque ya han aparecido en el mercado dispositivos con resoluciones mucho mayores (la denominada resolución 4K), pensamos que tomar como resolución objetivo para esta tesis la resolución Full-HD es una decisión más práctica y realista, debido sobre todo a la falta de contenido 4K que hay disponible actualmente.

- **Codificación: H.264/AVC**. El formato de vídeo H.264 se está imponiendo como la solución de facto para la codificación y decodificación de vídeo en Internet. La empresa Zencoder ofrece algunas estadísticas sobre la utilización de codecs de vídeo y audio en [Zencoder, 2010]. Zencoder proporciona servicios de codificación de audio y vídeo en la nube mediante el paradigma "software as a service", por lo que manejan datos tanto de los formatos de entrada como de los formatos de salida de los vídeos que ellos procesan. En estos datos se puede ver la supremacía de H.264 como el codec más utilizado del momento.

- **Sistema de transmisión sin errores**. Como se vio en el capítulo 3, el efecto que introduce la red se va a considerar como una degradación de la calidad audiovisual, por lo que el modelo de vídeo que se desarrolla en esta sección supone que la red no introduce degradación alguna en la calidad. Así pues, este modelo solo analiza el efecto que tiene en la calidad el proceso de codificación de vídeo.

4.2. Revisión del estado del arte

En esta sección se lleva a cabo una revisión de los trabajos más destacados en el ámbito de la estimación de la calidad de vídeo mediante modelos objetivos o métricas objetivas de calidad. Estas métricas objetivas de calidad son algoritmos diseñados para caracterizar la calidad de una secuencia de vídeo y para predecir o estimar la valoración de un usuario. Más concretamente, las métricas objetivas de calidad son las herramientas que permiten llevar a cabo evaluaciones objetivas de calidad, las cuales tienen como objetivos:

- Definir un método fiable para la estimación de MOS, es decir, la predicción ofrecida por las métricas objetivas de calidad debe estar fuertemente relacionada con la valoración de los usuarios.

- Definir un método repetible para la estimación de MOS, es decir, dos valoraciones de la misma métrica objetiva de calidad sobre las mismas secuencias de vídeo deberían proporcionar los mismos resultados.

Una posible clasificación de las métricas objetivas de calidad se puede realizar en función de los inputs que necesita la métrica para generar la valoración de calidad. Mediante este criterio se pueden distinguir los siguientes tipos de métricas o modelos:

- Métricas de datos: estas métricas miden la fidelidad de la señal sin tener en cuenta su contenido. A este grupo pertenecen métricas como Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), las cuales no tienen en cuenta la importancia visual de los píxeles de cada imagen.

- Métricas de imagen: a esta categoría pertenecen aquellas métricas que consideran la información visual que contiene el vídeo. Más concretamente, es habitual que este tipo de métricas tengan en cuenta el efecto de las distorsiones y el efecto del contenido en la calidad percibida. Se suelen basar en modelos derivados del sistema visual humano o bien en la extracción de ciertas características o artefactos de la secuencia de vídeo.

- Métricas de paquete o de flujo de bits: estos modelos se basan en extraer información directamente de las cabeceras y del flujo de bit codificado de vídeo, por lo que no necesitan decodificar la señal de vídeo para generar la predicción de calidad. Tienen la ventaja de necesitar poca cantidad de información para generar la predicción, por lo que la velocidad de procesamiento es alta. Por el contrario, la propia naturaleza de este tipo de métricas requiere que éstas estén específicamente diseñadas para codecs y protocolos de red concretos.

- Métricas híbridas: esta categoría engloba aquellas métricas que aplican dos o más enfoques de los comentados hasta el momento.

Otro enfoque que permite clasificar las métricas objetivas de calidad se basa en la cantidad de información de referencia que necesitan. Así pues, se puede distinguir entre:

- Métricas de referencia completa, FR: estas métricas miden la degradación que presenta una secuencia de vídeo con respecto a una secuencia de vídeo de referencia (sin degradaciones). Generalmente requieren que ambas secuencias estén alineadas espacial y temporalmente, además de estar calibradas en cuanto a nivel de color, con el objetivo de poder comparar los píxeles de cada trama en ambas secuencias de vídeo.

- Métricas de referencia nula o sin referencia, NR: las métricas sin referencia solo tienen en cuenta la secuencia de vídeo degradada, por lo que no necesitan referencia ni ningún tipo de calibrado. El principal desafío al que se enfrentan este tipo de modelos es la distinción de distorsiones dentro del contenido del vídeo.

- Métricas de referencia reducida, Reduced Reference (RR): representan el punto medio entre las métricas FR y RR en cuanto a cantidad de información de referencia que necesitan. Las métricas RR extraen un conjunto de parámetros de la señal de referencia y realizan sus predicciones en torno a ellas.

Como se puede ver, el ámbito de aplicación marca la naturaleza de la métrica objetiva de calidad a utilizar. Por ejemplo, para llevar a cabo comparaciones entre codecs, es recomendable utilizar métricas FR, mientras que para la monitorización online de la QoE las métricas NR o RR son las más deseables.

A continuación, se presentan los trabajos más destacados en el ámbito de las métricas objetivas de calidad de vídeo. En primer lugar se hace un repaso a las aportaciones de diferentes organismos de estandarización y foros de la industria. Tras esto se presentan y se analizan varios artículos científicos que contienen diversas contribuciones de interés para esta tesis.

4.2.1. Proyectos Video Quality Expert Group

El Video Quality Experts Group (VQEG) fue fundado por un grupo de miembros de ITU-T e ITU-R en 1997. Este grupo está formado por expertos en evaluación de calidad de vídeo, tanto del ámbito académico como industrial. El principal objetivo del VQEG es contribuir a la rama de la evaluación de calidad de vídeo mediante la validación de métricas objetivas y el desarrollo de nuevos métodos de evaluación subjetiva [Brunnstrom et al., 2009].

Para llevar a cabo la validación de métricas objetivas, el VQEG genera bases de datos de secuencias de vídeo de prueba y lleva a cabo experimentos de evaluación subjetiva de calidad. Las secuencias de vídeo de prueba no se proporcionan a los desarrolladores de métricas de calidad, ya que el proceso de evaluación consiste en obtener las predicciones de calidad que cada métrica genera para cada secuencia de vídeo, y compararlas con los resultados de las evaluaciones subjetivas. La valoración que el VQEG ofrece para cada métrica o modelo se basa en el rendimiento que obtienen las predicciones generadas en base a criterios estadísticos.

Las actividades del VQEG se organizan en proyectos, cada uno de ellos orientado a evaluar métricas de calidad que comparten un conjunto de características. A continuación se proporciona una breve descripción de los proyectos que el VQEG ya ha concluido.

FRTV Phase I Este es el primer proyecto del VQEG, el cual fue completado en junio de 2000. Este proyecto se centró en métricas FR y en secuencias de definición estándar, principalmente codificadas en MPEG-2 con diferentes perfiles y parámetros.

Se utilizaron 20 secuencias de vídeo (cada una de ellas codificada con distintos parámetros), las cuales fueron evaluadas de manera subjetiva utilizando el método Double Stimulus Continuous Quality Scale (DSCQS). Los resultados de este test demostraron que todas las métricas evaluadas eran estadísticamente equivalentes al PSNR, por lo que ninguno de los modelos pudo ser recomendado para su utilización.

FRTV Phase II El segundo proyecto (FRTV Phase II) fue completado en agosto de 2003. Este proyecto es una ampliación del proyecto anterior, en el que se aumentó el número de secuencias de vídeo y el número de degradaciones aplicadas a cada secuencia. En este caso, los resultados fueron más positivos que en el proyecto anterior, ya que la mejor métrica obtuvo una correlación del 94 % con MOS, superando claramente al PSNR, cuya correlación se sitúa en torno al 70 %. Los cuatro algoritmos con mejor valoración en este proyecto se estandarizaron en ITU-T J.144 [ITU, 2004c] y en ITU-R BT.1683 [ITU, 2004a]. Estos modelos son los siguientes:

- British Telecom (United Kingdom, VQEG Proponent D), Anexo A de ITU-T J. 144.
- Yonsei University / SK Telecom / Radio Research Laboratory (Republic of Korea, VQEG Proponent E), Anexo B de ITU-T J.144.
- CPqD (Federative Republic of Brazil, VQEG Proponent F), Anexo C de ITU-T J.144.
- National Telecommunications and Information Administration (NTIA) (United States of America, VQEG Proponent H), Anexo D de ITU-T J.144.

En términos absolutos, el modelo de NTIA obtuvo la mayor correlación con respecto a las valoraciones subjetivas en las 525 secuencias probadas. Como se verá más adelante, **el modelo Video Quality Model (VQM) del NTIA, es de especial relevancia para esta tesis**.

Multimedia Phase I Completado en septiembre de 2008, este proyecto se centra en la evaluación de la calidad multimedia (o audiovisual) en secuencias de vídeo con bajas tasas de bit de codificación y tamaño de trama reducido (resolucion QCIF, CIF y VGA). Se evaluaron modelos de todo tipo (FR, RR y NR), dando lugar a las recomendaciones ITU-T J.247 [ITU, 2008d], que define cuatro modelos FR; y a la recomendación ITU-T J.246 [ITU, 2008c], que define tres modelos RR. Como se puede ver, ningún modelo NR obtuvo el rendimiento necesario para ser incluido en las recomendaciones de ITU.

Los modelos FR recomendados son los siguientes:

- NTT (Japan, VQEG Proponent A), Anexo A de ITU-T J.247.

- OPTICOM (Germany, VQEG Proponent B), Anexo B de ITU-T J.247.
- Psytechnics (United Kingdom, VQEG Proponent C), Anexo C de ITU-T J.247.
- Yonsei University (Republic of Korea, VQEG Proponent D), Anexo D de ITU-T J.247.

Los modelos RR recomendados son variaciones con distintos ancho de banda de referencia de un modelo RR de la universidad de Yonsei (Korea).

- Yonsei RR10k
- Yonsei RR64k
- Yonsei RR128k

RRNR-TV El objetivo de este proyecto fue evaluar modelos NR y RR para secuencias de televisión de definición estándar (525 y 625 líneas), codificadas en MPEG-2 y H.264. Las evaluaciones subjetivas fueron realizadas utilizando el método ACR. Este proyecto fue completado en junio de 2009 y como resultado del mismo ITU estandarizó varios modelos RR en ITU-T J.249 [ITU, 2010a]:

- Model-A 15k, Yonsei University, HDSP Laboratory, Anexo A de ITU-T J.249
- Model-A 80k, Yonsei University, HDSP Laboratory, Anexo A de ITU-T J.249
- Model-A 256k, Yonsei University, HDSP Laboratory, Anexo A de ITU-T J.249
- Model-C 80k, NTIA, Anexo C de ITU-T J.249
- Model-B 80k (525-line only), NEC, Anexo B de ITU-T J.249
- Model-B 256k (525-line only), NEC, Anexo B de ITU-T J.249

Además, ITU estandarizó la implementación de PSNR que se utilizó en este proyecto en la recomendación ITU-T J.340 [ITU, 2010b].

HDTV Phase I Completado en junio de 2010, en este proyecto el VQEG validó diversos modelos de calidad de vídeo (FR, RR y NR) para televisión de alta definición High Definition Television (HDTV). Todas las secuencias de vídeo utilizadas tenían una resolución de 1920 x 1080, aunque se hicieron pruebas con versiones escaladas a 720p. Los codecs utilizados fueron MPEG-2 y H.264, con tasas de bit de entre 1 y 30 Mbps. Cada secuencia se procesó utilizando diferentes Hypothetical Reference Circuit (HRC), los cuales introdujeron artefactos de compresión y errores de transmisión. Los tests subjetivos se realizaron utilizando el método ACR con referencia oculta.

Los modelos NR fueron desestimados, mientras que algunos de los modelos FR y RR fueron estandarizados por ITU. Más concretamente, en ITU-T J.341 [ITU, 2011b] se define el modelo FR que mejor puntuación obtuvo en las evaluaciones: VQuad-HD, desarrollado por SwissQual (Suiza). En ITU-T J.342 [ITU, 2011c] se define el modelo RR que mejor puntuación obtuvo en las evaluaciones: Yonsei-HDRR, desarrollado por la universidad de Yonsei (Korea), con versiones de 56k, 128k y 256k de ancho de banda de referencia.

Proyectos en curso Además de estos proyectos, el VQEG está desarrollando un nuevo proyecto denominado **AVHD (Audiovisual HD Quality)**, donde se evaluarán nuevas métricas de calidad de vídeo y de calidad audiovisual. Este proyecto surge como la fusión de dos proyectos, HDTV2 y Multimedia 2, razón por la que se evaluarán modelos de audiovisuales y de vídeo. Los resultados de este proyecto no están disponibles en la fecha de escritura de esta tesis.

Otros proyectos en curso son: 3DTV, HDR (High Dynamic Range Video), Hybrid Perceptual/Bitstream, JEG-Hybrid, MOAVI (Monitoring of Audio Visual Quality by Key Indicators), Quality Recognition Tasks (QART), RICE (Real-Time Interactive Communications Evaluation) y Ultra HD.

Como se puede ver, el VQEG contribuye enormemente al avance de la evaluación (tanto objetiva como subjetiva) de calidad de vídeo. Su labor de evaluación independiente permite la generación de estándares, además de proporcionar herramientas que mejoran y facilitan el desarrollo de nuevas métricas objetivas de evaluación de calidad de vídeo.

4.2.2. Recomendaciones International Telecommunication Union (ITU)

En materia de métricas objetivas de calidad, ITU y el VQEG colaboran estrechamente. El VQEG reporta los resultados de sus proyectos a los grupos de estudio 9 (Broadband cable and TV) y 12 (Performance, QoS and QoE) de ITU-T y al grupo de estudio 6 (Broadcasting service) de ITU-T.

A continuación se describen con más detalle las recomendaciones ITU que han surgido como resultado de las evaluaciones realizadas por parte del VQEG.

4.2.2.1. ITU-T J.144 e ITU-R BT.1683

Modelo General VQM NTIA Como se comentó anteriormente, ITU-T J.144 [ITU, 2004c] e ITU-R BT.1683 [ITU, 2004a] describen los cuatro modelos con mayor rendimiento en el test FRTV Phase II de VQEG. De estos cuatro modelos, el que obtuvo mayor puntuación fue el modelo VQM general de NTIA, el cual se describe a continuación.

El modelo VQM de NTIA, descrito también en [Pinson and Wolf, 2004], se diseñó como una métrica objetiva de calidad de vídeo de propósito general, aplicable a sistemas con un amplio rango de calidad y tasa de bit. En el diseño de este modelo se llevaron a cabo numerosas evaluaciones subjetivas de calidad con el objetivo de analizar el rendimiento de éste, antes de ser presentado al VQEG.

En la figura 4.1 se muestra el proceso necesario para obtener VQM según NTIA.

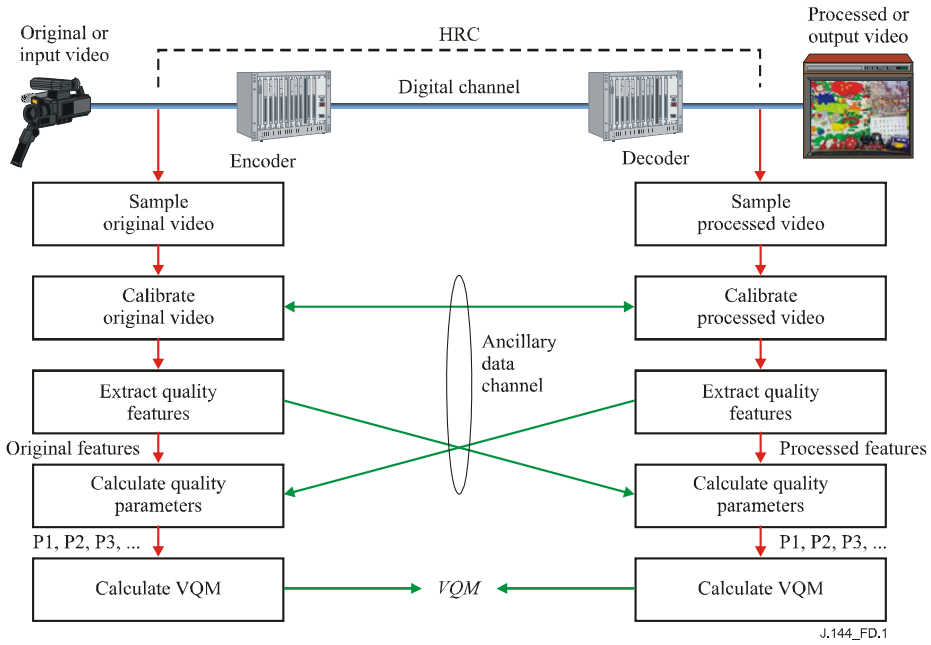


Figura 4.1: Proceso de cálculo de VQM. [ITU, 2004c]

A grandes rasgos, se puede ver que aunque este modelo ha sido presentado como un modelo de referencia completa, y puede usarse como tal, realmente es un modelo de referencia reducida, ya que solo se extraen ciertas características de la señal de referencia que pueden ser transmitidas mediante un canal de comunicación auxiliar, haciendo posible la evaluación de la calidad en tiempo casi real. Más concretamente, se extraen un conjunto de parámetros de ciertas regiones espacio-temporales (S-T regions), las cuales requieren, según los autores un 9,3 % del ancho de banda de la secuencia sin comprimir. A este ancho de banda, hay que añadir un 4,7 % asociado a las técnicas de calibración que el modelo requiere.

Como se puede ver en la figura, la primera fase es un proceso de muestreo, cuyo objetivo es representar digitalmente una señal de vídeo si ésta fuese analógica.

La fase de calibrado incluye alineación espacial y una estimación de la región válida del vídeo (para sistemas de televisión donde ciertas líneas se pierden). La alineación espacial es un proceso mediante el cual se determina el desplazamiento horizontal y

vertical del vídeo procesado o degradado con respecto al vídeo original, y una vez determinado se cancela dicho desplazamiento. Esta fase incluye también otros cálculos, como el de la ganancia y el offset en los valores de luminancia, además del cálculo de la alineación temporal. De especial interés resulta la estimación de la región válida de procesado o PVR (Processed Valid Region), la cual elimina ciertos bordes de la imagen que pueden no contener información válida. Este comportamiento puede producirse en secuencias de vídeo que hayan sido muestreadas de acuerdo a la recomendación ITU-R BT.601 [ITU, 2012b] o en sistemas de compresión que reduzcan el área de la imagen para reducir la información a transmitir.

Una vez calibradas las secuencias de vídeo, se extraen un conjunto de "quality features", las cuales se definen como una cantidad de información asociada con, o extraída de, una sub-región espacio-temporal válida de una de las secuencias de vídeo (original o degradada). Desde un punto de vista de alto nivel, todas las "quality features" se extraen siguiendo el mismo procedimiento. En primer lugar se aplica un filtro perceptual al flujo de vídeo para realzar alguna propiedad de la calidad percibida del vídeo (por ejemplo información de bordes de la imagen). Tras este filtrado se extrae un valor concreto para cada sub-región espacio-temporal aplicando alguna función matemática (por ejemplo la desviación típica). Por último, se aplica un filtro de perceptibilidad a los valores extraídos.

Mediante la comparación de "quality features" extraídas de la secuencia original y de la secuencia degradada se obtiene lo que en este modelo se denominan parámetros de calidad o "quality parameters", los cuales son indicadores de los cambios perceptuales que se han producido en la calidad de vídeo. En primer lugar, se realiza una comparación entre regiones espacio-temporales para la secuencia original y la secuencia degradada. Después, los resultados de estas comparaciones a nivel de sub-región se agregan utilizando alguna función de pooling, generando un valor individual para la secuencia completa de vídeo, la cual se supone de unos 8 o 10 segundos de duración.

En la especificación del modelo se incluyen siete "quality parameters". Cuatro de ellos basados en "quality features" extraídas de gradientes espaciales de la componente de luminancia, dos parámetros se basan en el vector formado por las componentes de crominancia y el último parámetro se basa en el producto de "quality features" que miden el contraste y la cantidad de movimiento (ambas extraídas de la componente de luminancia). En concreto, estos siete parámetros son los siguientes:

- si_loss: detecta la pérdida o el descenso en la información espacial (blurring o difuminado).

- hv_loss: detecta la transformación de bordes horizontales y verticales a bordes diagonales.

- hv_gain: es el complementario al parámetro anterior, ya que detecta la transformación de bordes diagonales a bordes horizontales y verticales (blockiness, tiling).

- chroma_spread: detecta cambios en la extensión de la distribución de muestras de color.

- si_gain: cuantifica las mejoras en la calidad que puedan resultar del afilado de bordes.

- ct_ati_gain: es el producto de un valor de contraste y una medida de información temporal, detectando errores en el movimiento de los bordes.

- chroma_extreme: es una variación de chroma_spread utilizando diferentes funciones de agregación espacio-temporal, la cual se utiliza para detectar degradaciones en la información de color asociadas a errores en la transmisión.

Por último, el modelo general VQM consiste en la siguiente combinación lineal de los siete "quality parameters" anteriores.

$$\begin{aligned} VQM = &-0,2097 \cdot si\_loss + 0,5969 \cdot hv\_loss + 0,2483 \cdot hv\_gain \\ &+0,0192 \cdot chroma\_spread - 2,3416 \cdot si\_gain \\ &+0,0431 \cdot ct\_ati\_gain + 0,0076 \cdot chroma\_extreme \end{aligned} \tag{4.1}$$

Se debe destacar que $si\_loss \leq 0$ y que el resto de parámetros son siempre iguales o mayores que 0. Así pues, si_gain es el único parámetro que puede disminuir el valor de VQM.

Como se deduce de la ecuación 4.1, VQM es una medida de degradación, por lo que la recomendación incluye una función de recorte para que no pueda alcanzar valores negativos (lo cual implicaría una mejora en la calidad por parte de la secuencia degradada). Se incluye también una expresión para permitir un valor máximo de 1,5 (para secuencias extremadamente degradadas), aunque los valores habituales de VQM van de 0 a 1.

$$VQM = \begin{cases} 0, & VQM \leq 0 \\ \frac{1,5 \cdot VQM}{0,5 + VQM}, & VQM > 1 \end{cases} \tag{4.2}$$

4.2.2.2. ITU-T J.247

Modelo NTT El modelo NTT se divide en tres módulos:

- Módulo de alineado de vídeo: lleva a cabo dos procesos, un proceso de macro-alineado que consiste en relacionar los píxeles entre la señal de referencia y la señal degradada tanto espacial como temporalmente, y un proceso de micro-alineado que relaciona tramas entre la señal de referencia y la señal degradada teniendo en cuenta tramas perdidas o duplicadas (freezing).

- Módulo de derivación de características (features) espacio-temporales: este módulo calcula un parámetro de degradación espacial y un parámetro de degradación temporal, utilizando las señales alineadas que proporciona el módulo anterior. El parámetro de degradación espacial se basa en cuatro sub-parámetros que evalúan la presencia de degradaciones como ruido, bordes espurios, distorsión de movimiento y otras distorsiones espaciales. El parámetro de degradación temporal, estima el efecto de la variación de la tasa de frames y de las congelaciones en la imagen.

- Módulo de estimación de calidad subjetiva de vídeo: este módulo se encarga de realizar la predicción de la calidad de vídeo, en términos de Difference Mean Opinion Score (DMOS).

Como se puede ver, este modelo sigue una estructura similar al modelo general de NTIA descrito anteriormente.

Modelo PEVQ (Opticom) El modelo PEVQ (Perceptual Evaluation of Video Quality), desarrollado por Opticom, sigue un proceso similar a los modelos analizados hasta el momento para obtener su predicción de la calidad.

En primer lugar, realiza el alineado entre la señal de referencia y la señal degradada. Después se realizan una serie de comparaciones, tanto a nivel de luminancia como a nivel de crominancia, entre las señales alineadas. Estas comparaciones dan lugar a cinco indicadores, basados en el sistema visual humano, los cuales se integran mediante funciones no lineales para obtener la predicción de MOS.

Modelo Psytechnics De manera análoga a los modelos anteriores, el modelo de Psytechnics sigue la misma arquitectura de tres bloques: alineado, extracción de parámetros, predicción de calidad. Para la extracción de parámetros, aplican un modelo de sistema visual humano, con el objetivo de identificar errores y artefactos visibles, como consecuencia del proceso de codificación y transmisión.

Modelo Yonsei FR Este modelo se basa en la observación de que el sistema visual humano es especialmente sensible a las degradaciones que se producen alrededor de los bordes de los objetos que aparecen en una imagen. Así pues, en primer lugar realizan

una detección de bordes en la señal original, la cual comparan con los bordes en la señal degradada en términos de MSE. A partir de este valor de MSE obtienen una medida del PSNR de los bordes o EPSNR (Edge PSNR), la cual combinan con otros dos parámetros (que miden el nivel de blurriness y de blockiness) para obtener la predicción de calidad de vídeo.

4.2.2.3. ITU-T J.246

Modelo Yonsei RR Este modelo es una variación del modelo FR de la universidad de Yonsei, definido en ITU-T J.247, adaptado a una configuración RR. Así pues, el algoritmo subyacente es el mismo (EPSNR), con la salvedad de que solo un conjunto de píxeles correspondientes a bordes de la señal de referencia son contemplados para llevar a cabo la comparación con la señal degradada. En función de la cantidad de información de la señal de referencia que se utilice, variará el ancho de banda necesario para transmitir dicha información, dando lugar a tres versiones del mismo modelo: Yonsei RR10k, Yonsei RR64k y Yonsei 128k. Sin embargo, con las resoluciones consideradas en este recomendación (CIF, QCIF y VGA), basta con un canal auxiliar de 10kbps para las resoluciones CIF y QCIF y de 30kbps para la resolución VGA.

4.2.2.4. ITU-T J.249

Modelo Yonsei RR La recomendación ITU-T J.249 [ITU, 2010a] define una nueva versión del modelo RR de la universidad de Yonsei. Como se describió anteriormente, el algoritmo en el que se basa este modelo es el EPSNR. En esta recomendación el modelo sigue el mismo enfoque, sin embargo, en la última fase del proceso de estimación de calidad, se incluye un conjunto de modificaciones al valor de EPSNR con el objetivo de cuantificar la congelación de tramas, secuencias con mucho movimiento, blurriness y blockiness. Como se puede ver, este modelo incluye algunas de las características del modelo FR de Yonsei que el anterior modelo RR no incluía.

Modelo NEC El modelo RR propuesto por NEC se basa en el concepto de "actividad", que se define como la media de la diferencia entre los valores absolutos de luminancia y la media de la luminancia para un bloque de tamaño dado. Como se puede ver, lo que NEC denomina actividad es una medida de dispersión de los valores de luminancia de un bloque de píxeles.

En base a este concepto, el modelo de NEC propone las siguientes fases:

1. Se calculan los valores de actividad para cada bloque de 16x16 píxeles de la secuencia de referencia. Esta información es la que se transmite al cliente para realizar la estimación de calidad.

2. Se calculan los valores de actividad correspondientes sobre la secuencia degradada.

3. Para cada bloque se calcula el error cuadrático entre los valores de actividad de la secuencia original y degradada.

4. Se aplican pesos al error de aquellos bloques en los que se detecte alto nivel de movimiento o cambios de escena.

5. Se calcula una estimación provisional de la calidad de vídeo como una suma ponderada de los errores de cada bloque (teniendo en cuenta los pesos anteriores).

6. Se refina la estimación anterior teniendo en cuenta otras degradaciones como el blockiness.

NTIA Fast Low Bandwidth VQM Como se vio anteriormente, el modelo desarrollado por NTIA y definido en [ITU, 2004c] e [ITU, 2004a] es realmente un modelo RR. Teniendo esto en cuenta, el modelo definido en ITU-T J.249 [ITU, 2010a] se puede ver como una versión con ancho de banda reducido del modelo general VQM de NTIA. Los parámetros que se calculan y el proceso de obtención de calidad es análogo al del modelo general, pero reduciendo el ancho de banda necesario para transmitir la información de referencia a valores comprendidos entre 12 y 14 kbps.

4.2.2.5. ITU-T J.342

Modelo Yonsei HDRR En ITU-T J.342 [ITU, 2011c] se define una nueva versión del modelo RR de la universidad de Yonsei. Esta nueva variación del modelo está orientada a contenido HD codificado tanto en H.264 como en MPEG-2, e introduce algunas mejoras en la estimación del blockiness e incluye un nuevo bloque para contabilizar los errores de transmisión. Sin embargo, sorprende que, aunque este modelo está recomendado por ITU su rendimiento es similar al obtenido por el PSNR.

4.2.2.6. ITU-T J.341

Modelo VQuad En ITU-T J.314 [ITU, 2011b] se presenta el modelo FR que mejor rendimiento mostró en los tests de calidad de vídeo HDTV del VQEG. Este modelo es conocido comercialmente como VQuad-HD y está desarrollado por SwissQual.

El modelo de predicción se basa en modelos cognitivos y psico-visuales para emular la percepción subjetiva. Más concretamente, la predicción de calidad se realiza siguiendo estos pasos, los cuales se muestran también en la figura 4.2:

1. Procesado inicial de las secuencias de vídeo: filtrado para reducir ruido y submuestreo de tramas.

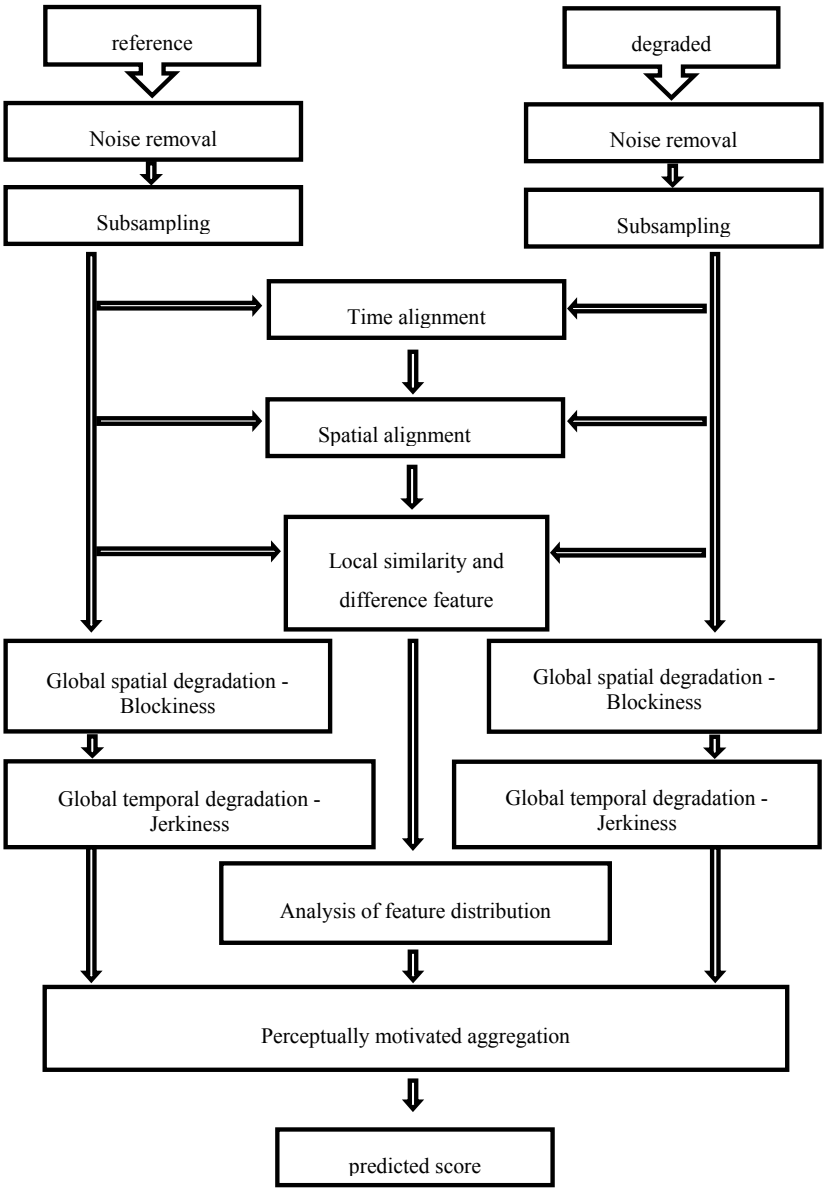


Figura 4.2: Proceso de cálculo de VQuadHD. [ITU, 2011b]

2. Alineado temporal de tramas entre la secuencia de referencia y la secuencia degradada.

3. Alineado espacial de tramas entre la secuencia de referencia y la secuencia degradada.

4. Cómputo de parámetros locales de calidad: medidas de similaridad y diferencias espaciales, inspiradas por los modelos de percepción visual humana.

5. Análisis de la distribución de los parámetros anteriores.

6. Cómputo de degradación espacial global, utilizando un módulo de "blockiness".

7. Cómputo de degradación temporal global, utilizando un módulo de "jerkiness" o "entrecortamiento" temporal. Este parámetro se calcula analizando la intensidad del movimiento y los tiempos en los que se ha visualizado cada trama del vídeo.

8. Estimación de calidad combinando los parámetros anteriores mediante una función de agregación no lineal.

4.2.3. Artículos científicos

En esta sección se presentan algunos de los artículos científicos más destacados en materia de métricas y modelos objetivos de calidad de vídeo.

4.2.3.1. Métricas sin referencia

En [Yang et al., 2005] se propone una métrica sin referencia para estimar la calidad de vídeo de secuencias que contienen escenas naturales.

Este modelo tiene en cuenta la dependencia temporal propia de las secuencias naturales y analiza la diferencia entre regiones de alta complejidad espacial en tramas sucesivas. Más concretamente, este trabajo explota la hipótesis de que el sistema visual humano espera que la secuencia de imágenes que forman un vídeo sea consistente. Así pues, si se producen cambios abruptos, inducidos por defectos de la codificación, esto conlleva una reducción en la calidad percibida. Esta idea lleva a los autores a utilizar medidas de diferencia entre tramas sucesivas para identificar dichos cambios, centrándose en regiones del vídeo con alta complejidad espacial que se mueven a lo largo de varias tramas. Además, estos valores de diferencias inter-trama, se ponderan utilizando una medida de la actividad o complejidad temporal de las tramas, dando más peso a aquellos bloques con menor movimiento, aplicando la hipótesis de que el sistema visual humano es más tolerante a errores en regiones con mucho movimiento.

El modelo fue entrenado utilizando las secuencias SRC18-SRC21 correspondientes al proyecto FR-TV Phase I del VQEG, y evaluado utilizando el resto de secuencias

de dicho proyecto. Se debe destacar que la propia concepción del modelo conlleva que ciertas secuencias no sean evaluadas correctamente: escenas no naturales y escenas con niveles de zoom cambiantes (lo cual dificulta la estimación de los vectores de movimiento entre tramas adyacentes). Estos problemas se ponen de manifiesto en las secuencias SRC3 y SRC8. Así pues, sin contar estas escenas, el modelo consigue un coeficiente de correlación de Pearson de 0,85 y un Root Mean Squared Error (RMSE) de 0,12 mientras que si se incluyen todas las secuencias en los cálculos, éstos resultan ser 0,65 y 0,22 respectivamente.

En [Farias and Mitra, 2005] se presenta una métrica de calidad sin referencia basada en la estimación del nivel de degradación asociado a tres artefactos: blockiness, blurriness y noisiness, para vídeo MPEG-2. Una vez obtenida la estimación del nivel de cada artefacto, estos se combinan utilizando un modelo lineal o un modelo de Minkowski. Aunque los resultados obtenidos son razonables (coeficiente de correlación de Pearson de 0,86), la poca variedad de secuencias de vídeo utilizadas (solo 6) plantea algunas dudas sobre la validez de dichos resultados.

[Ries et al., 2007] proponen un modelo de estimación de calidad sin referencia para secuencias de vídeo de baja resolución y orientado a dispositivos móviles. Más concretamente, se centra en resoluciones QCIF a un máximo de 105 kbit/s y resoluciones CIF y SIF a un máximo de 200 kbit/s.

El funcionamiento del modelo se basa en dos fases:

1. Clasificación del tipo de contenido: definen 5 clases de secuencias en función del contenido de las mismas (noticias, fútbol, dibujos animados, escenas panorámicas y otras) y llevan a cabo una clasificación de secuencias utilizando la señal original. Es importante destacar este punto, ya que aunque el modelo se puede considerar sin referencia, la clasificación del tipo de secuencia la realizan sobre la secuencia original.

2. Estimación de MOS sobre la secuencia codificada: proponen un modelo de estimación de MOS utilizando como parámetros la tasa de codificación, la tasa de frames y el tipo de secuencia (resultado del paso anterior). El modelo matemático que proponen se rige por la siguiente expresión:

$$MOS = A + B \cdot BR + \frac{C}{BR} + D \cdot FR + \frac{E}{FR} \tag{4.3}$$

   Finalmente, para cada tipo de secuencia de vídeo, llevan a cabo un ajuste de los parámetros del modelo obteniendo una correlación entre la predicción del modelo y la MOS medida en experimentos subjetivos que va desde el 99 % para el caso mejor, hasta el 75 % en el caso peor.

Es interesante la clasificación que los autores de este artículo llevan a cabo en cuanto

a tipos de secuencia de vídeo, ya que asienta la idea de que el contenido de la secuencia es fundamental a la hora de estimar su calidad. Esta es una idea ampliamente utilizada en la literatura actual, y será también aplicada en esta tesis.

Por otro lado, en cuanto al ajuste del modelo (como se verá más adelante) el factor correspondiente a la tasa de codificación (que en este artículo sigue una variación de la forma $x + \frac{1}{x}$) no corresponde a las medidas realizadas en esta tesis. Esta discrepancia puede deberse al conjunto de resoluciones y tasas de bits utilizadas en este artículo, las cuales son varios órdenes de magnitud menores que las utilizadas en esta tesis.

En [Naccari et al., 2009] se define el algoritmo NORM (NO-Reference video quality Monitoring), diseñado para cuantificar la degradación en la calidad de los errores de canal sobre vídeos codificados con H.264/AVC. NORM analiza la distorsión introducida por las técnicas de ocultación de errores espaciales y temporales además del efecto de la compensación de movimiento. Con esta información, el algoritmo propuesto genera una estimación del MSE a nivel de macrobloque. Se debe destacar que el nombre del algoritmo puede dar lugar a confusión, al incluir el término "no reference". Lo que realmente proponen los autores es introducir la estimación del MSE (sin utilizar información de referencia) en un modelo de referencia reducida para obtener un valor de SSIM, el cual correla con la valoración subjetiva de los usuarios.

El modelo propuesto en [Keimel et al., 2009] sigue un enfoque similar al modelo de Farias y Mitra [Farias and Mitra, 2005], analizado anteriormente. En concreto, dicho modelo estima la calidad percibida en vídeos H.264/AVC y SVC de alta definición mediante el análisis (realizado sobre una región central del vídeo) de cuatro parámetros: blockiness, blurriness, nivel de actividad (cantidad de información espacial) y predictibilidad (nivel de información temporal). El diseño del modelo se basa en el ajuste numérico de una combinación lineal de los parámetros anteriores. Uno de los puntos "débiles" de este modelo es que está entrenado unicamente con siete secuencias de vídeo, por lo que los resultados no son del todo fiables. Para intentar suplir este problema los autores incluyen un proceso de corrección que se basa en generar una nueva secuencia de vídeo, degradando la señal recibida, con el objetivo de analizar, para cada secuencia de vídeo concreta, cómo de sensible es con respecto a la métrica de calidad y corregir, si fuera necesario, la valoración de calidad basada en los cuatro parámetros iniciales.

Otro enfoque similar al anterior se propone en [Kawano et al., 2010], modelo que utiliza como parámetros de calidad el nivel de blurriness y blockiness de la señal de vídeo degradada para generar la estimación de QoE.

El modelo propuesto en [Brandao and Queluz, 2010] sigue la estructura clásica de modelo sin referencia: estimación de parámetros de calidad o degradación y ponderación y agregación de los mismos para obtener la predicción de calidad. Sin embargo, este trabajo propone nuevas ideas en cuanto a la estimación de los parámetros de ca-

lidad para vídeo H.264/AVC. A diferencia de los modelos anteriores, en los que las estimaciones se realizaban sobre los valores de luminancia y crominancia de los píxeles de cada trama del vídeo, Brandao y Queluz proponen llevar esta estimación de parámetros al dominio de la frecuencia. Así pues, la estimación de los parámetros de calidad (lo que en el artículo denominan estimación del error) se basa en información sobre los coeficientes de la DCT.

Aunque existen trabajos anteriores cuyo objetivo es predecir el valor del PSNR mediante el análisis de los coeficientes de la DCT sin utilizar referencia, el trabajo de Brandao y Queluz es el primero en dar el salto a estimaciones de QoE en el dominio de la DCT.

Otra propuesta parecida se puede encontrar en [Saad and Bovik, 2012], la cual define el modelo BLIINDS. Este modelo de estimación de calidad de vídeo sin referencia se basa en un modelo estadístico de los coeficientes de la DCT en escenas naturales y en un modelo temporal que analiza la coherencia del movimiento.

Este modelo se basa en la observación de que las escenas naturales comparten una serie de estadísticos bastante fiables y regulares. Partiendo de esta base, la desviación con respecto a estas estadísticas será consecuencia de la degradación que ha sufrido el vídeo y se podrá estimar una valoración de la calidad percibida en base a dicha desviación. En concreto, BLIINDS analiza la distribución de los coeficientes de la DCT aplicada a la diferencia entre tramas. Por otro lado, BLIINDS también caracteriza el tipo de movimiento que se produce en la secuencia de vídeo.

En [Joskowicz et al., 2009] los autores proponen modelar la calidad percibida en secuencias de vídeo codificadas en MPEG-2 y en H.264 con resoluciones VGA, CIF y QCIF, en términos de DMOS mediante una sencilla fórmula matemática, función de la tasa de bit de codificación. Este modelo utiliza como métrica de referencia el modelo general VQM de NTIA y define la siguiente relación:

$$DMOS = \frac{m}{k \cdot (a \cdot bitrate)^n} \tag{4.4}$$

En la ecuación 4.4, el parámetro k depende del códec utilizado. Los autores proponen los siguientes valores:

$$k = \begin{cases} 1, & \text{MPEG-2} \\ l + d \cdot e^{-b \cdot a \cdot bitrate}, & \text{H.264} \end{cases} \tag{4.5}$$

Los parámetros m y n, por su parte, se obtienen mediante un proceso de ajuste, donde los autores clasifican cualitativamente las secuencias de vídeo en función de la cantidad de movimiento. Los resultados de este ajuste se muestran en la tabla 4.1.

En [Pérez et al., 2011] se propone un enfoque híbrido entre el mundo de la calidad de

Tabla 4.1: Parámetros de ajuste del modelo Joskowicz et al

Tipo de secuencia	**m óptimo**	**n óptimo**	**MSE**
Poco movimiento	0,192	0,992	0,0264
Movimiento moderado	0,368	0,956	0,0346
Mucho movimiento	0,536	0,894	0,0616

servicio QoS y la calidad percibida QoE. Los autores afirman que las métricas objetivas de calidad actuales no son directamente aplicables a la monitorización continua de la calidad percibida, por lo que intentan encontrar un punto intermedio entre la objetividad y las ventajas en cuanto a facilidad de monitorización que tienen los parámetros de QoS y la correlación existente entre la calidad percibida por los usuarios y las métricas objetivas de calidad de vídeo. Para llevar a cabo esta tarea, la estrategia que siguieron los autores fue partir de la medida del Media Delivery Index (MDI), estándar de facto en la medida de QoS, que combina parámetros de red como la tasa de pérdidas y el jitter, fácilmente medibles en diversos puntos de la red. Como el MDI no puede ser utilizado directamente como medida de QoE, los autores intentan refinar esta métrica para acercarla a los resultados que obtienen las métricas objetivas de calidad de vídeo. En este contexto, los autores proponen una arquitectura denominada QuEM (Qualitative Experience Monitoring) la cual tiene como objetivo detectar la intensidad y la duración de diferentes degradaciones y artefactos que pueden aparecer en el streaming de vídeo.

En [Argyropoulos et al., 2011] se propone un enfoque similar al de [Pérez et al., 2011]. En este caso, el modelo que proponen los autores intenta estimar el impacto que tiene la tasa de pérdidas de paquetes en la percepción de la calidad del vídeo. Para ello, el algoritmo propuesto extrae ciertos parámetros del flujo de bit recibido y determina el efecto o la visibilidad de cada evento de pérdidas mediante la clasificación de los parámetros del flujo de bit usando una SVM (Support Vector Machine). Finalmente, el nivel de visibilidad de los eventos de pérdidas se mapea a un nivel de calidad percibida.

En [Leister et al., 2011] se propone un modelo basado en degradaciones para estimar la calidad percibida por un usuario teniendo en cuenta diversas fases y procesos de la cadena de distribución de vídeo. Este algoritmo se basa en el modelo E y propone la siguiente expresión:

$$Q = Q_0 \cdot \prod_{i \in \{E,S,N,U,V,A\}} M_i \tag{4.6}$$

Los factores M_i son factores de degradación, por lo tanto $M_i \leq 1$ excepto en el caso de M_A, que de manera análoga al modelo E, representa un factor de conveniencia, lo cual implica $M_A \geq 1$.

Cada uno de estos factores de degradación hace referencia a una parte de la cadena de distribución de vídeo:

- M_E: influencia del proceso de codificación. Depende del tipo de codec, los parámetros de codificación, etc.
- M_S: influencia del servidor de streaming. Depende del protocolo utilizado, la implementación del servidor, etc.
- M_N: influencia de la red. Depende del retardo, jitter, tasa de pérdidas, etc.
- M_U: influencia del equipamiento de usuario. Depende del tipo de hardware y software utilizado.
- M_V: influencia de las condiciones de visionado.
- M_A: factor de conveniencia para tener en cuenta la aceptación de ciertas degradaciones en función del tipo de contenido.

Del conjunto de factores de degradación, los autores desestiman algunos en su modelo. En concreto, proponen $M_U = 1$, $M_V = 1$ y $M_A = 1$. Además, combinan el efecto de M_E y M_S en un único factor de degradación $M_{E,S}$. En definitiva, el modelo se basa en analizar el efecto de la codificación (en términos de tasa de bit de codificación) y de la red (en términos de tasa de bit, tasa de pérdidas y retardo). Para ajustar la expresión de los factores de degradación los autores llevaron a cabo varios experimentos subjetivos. La relación encontrada entre la tasa de bit de codificación y la calidad fue logarítmica, mientras que la relación entre los parámetros de red considerados y la calidad no fue matemáticamente definida.

En [de la Cruz Ramos et al., 2012] se propone un modelo que comparte varias características con otras propuestas anteriores. En primer lugar, trata de estimar el valor de VQM según el modelo general de NTIA. Para ello, se basa en parámetros de red y de codificación, en concreto, en la tasa de bit de codificación y en la tasa de pérdidas del canal. El modelo propuesto por los autores se presenta en la ecuación 4.7.

$$VQM = VQM_C + VQM_L \tag{4.7}$$

En este modelo, VQM_C es la contribución del proceso de codificación al valor de VQM, mientras que VQM_L es la contribución de las pérdidas de paquete al valor de VQM. Estas dos componentes se pueden ajustar matemáticamente mediante las ecuaciones 4.8 y 4.9, donde VCR es la tasa de codificación, VCR_{REF} es la tasa de codificación de referencia (1Mbps), VQM_{REF} es el valor de VQM a la tasa de codificación de referencia, PLR es la tasa de pérdidas de paquete y PLR_1 es el valor de PLR para el cual $VQM = 1$.

$$VQM_C = VQM_{REF} \cdot \left(\frac{VCR}{VCR_{REF}}\right)^{-K_C} \tag{4.8}$$

$$VQM_L = (1 - VQM_C) \cdot \left(\frac{PLR}{PLR_1}\right)^{K_L} \tag{4.9}$$

Los parámetros de este modelo dependen del codec, de los parámetros de codificación y de las características de las secuencias de vídeo (complejidad espacial y temporal).

El enfoque seguido por los autores se basa en clasificar las secuencias de vídeo en función de dos métricas: Average Spatial Information (ASI) y Average Temporal Information (ATI), que evalúan la complejidad del contenido de las secuencias. Estas métricas derivan de las medidas de información espacial y temporal SI y TI, definidas en [Webster et al., 1993], y tienen como objetivo suplir la sensibilidad que las métricas originales SI y TI tienen en cuanto a valores excepcionalmente grandes en tramas individuales.

Así pues, los autores proponen utilizar los valores de ASI y ATI de una secuencia determinada como índices de una tabla precomputada con los parámetros del modelo, utilizando interpolación lineal en el caso de que los valores concretos de ASI y ATI de la secuencia no se encontraran en la tabla.

En [Hernando et al., 2013] se aborda el problema de la estimación de QoE en secuencias de vídeo codificadas en MPEG-2 mediante el análisis de las pérdidas de tramas MPEG. Los pasos que los autores han aplicado para desarrollar el modelo son los siguientes:

1. Generar una base de datos de vídeos con diferentes tasas de pérdida de paquetes.
2. Medir VQM según el modelo general de NTIA para cada secuencia de la base de datos.
3. Buscar una relación entre los valores de QoE estimados y la tasa de pérdidas de paquetes, incluyendo la influencia del tipo de trama MPEG afectada por las pérdidas.

Como resultado de este análisis, los autores proponen un modelo lineal como se puede ver a continuación:

$$MOS(I_{loss}, B_{loss}, P_{loss}) = 4,9030 - 1,0823 \cdot I_{loss} - 3,2792 \cdot B_{loss} - 3,2323 \cdot P_{loss} \tag{4.10}$$

Los autores plantean que aquellas secuencias con menor MOS tienen una mayor tasa de pérdidas de tramas I. Además, cuando se pierde información de la cabecera de la trama, dicha trama no puede decodificarse, por lo que debe tenerse en cuenta en el modelo.

Este modelo presenta un error de $0,113$ con respecto al modelo general de VQM, lo cual resulta en un error absoluto de $0,3367$ ($8,4\,\%$ en escala MOS) y un coeficiente de correlación cuadrático de Pearson $R^2 = 0,7575$.

4.2.3.2. Métricas de referencia completa

En [Wang et al., 2004] se propone un modelo de referencia completa para la estimación de la calidad de vídeo siguiendo un enfoque basado en la distorsión estructural de la imagen. Partiendo de la afirmación de que las imágenes naturales suelen tener una cierta estructura, la cual extrae el sistema visual humano para "entender" qué objetos están presentes en una imagen, este algoritmo propone un método directo para comparar la estructura de la señal original y la señal degradada.

La estimación de la degradación estructural se realiza mediante la métrica SSIM (Structural Similarity), la cual se define según la ecuación 4.11.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{4.11}$$

Los autores de este modelo aplican SSIM en tres niveles diferentes:

- A nivel local: seleccionan aleatoriamente un conjunto de bloques de tamaño 8x8 píxeles sobre los que calculan SSIM para cada una de las componentes de de luminancia y crominancia, obteniendo un valor agregado de SSIM para cada uno de estos bloques aplicando una suma ponderada, la cual da más peso a la componente de luminancia.

- A nivel de trama: se agregan los valores de SSIM de cada bloque de una trama aplicando unos pesos a cada bloque.

- A nivel de secuencia: se agregan los valores de calidad a nivel de trama aplicando pesos a cada trama.

Los pesos que proponen los autores se basan en las siguientes ideas:

- Las regiones oscuras atraen menos la atención que las zonas con más brillo, por lo que a las regiones oscuras se les asigna un peso menor.

- En escenas con mucho movimiento, ciertas degradaciones pueden pasar desapercibidas, por lo que a las regiones con mucha información temporal se les asigna un peso menor.

El algoritmo ha sido evaluado utilizando la base de datos de vídeos del experimento FR-TV Phase I del VQEG.

En [Wolf and Pinson, 2007] se lleva a cabo un experimento para evaluar el rendimiento del modelo general VQM de NTIA al ser aplicado a secuencias de vídeo de alta definición. En dicho experimento, se evaluaron 16 secuencias de vídeo 1920x1080i las cuales fueron codificadas con diferentes codecs a diferentes tasas de bit (de 2 Mbits/s a

18 Mbits/s), cuyas valoraciones de calidad subjetiva fueron comparadas con el resultado obtenido por el modelo objetivo. El coeficiente de correlación de Pearson obtenido fue de 0,84, mientras que la raíz del MSE fue de 9,7 (en una escala de 0 a 100).

En [Okamoto et al., 2009] evalúan el rendimiento del modelo de NTT, definido en [ITU, 2008d], al ser aplicado a secuencias de vídeo de alta definición. El resultado de esta evaluación es que el modelo de NTT no puede aplicarse directamente a secuencias HD, ya que aunque la correlación entre las valoraciones subjetivas y los resultados generados por el modelo tienen una buena correlación (0,87), la nube de puntos que relaciona ambas variables no sigue una tendencia lineal clara. Los autores proponen una modificación del modelo basada en combinar cada uno de los parámetros del mismo mediante medidas difusas, en vez de utilizar sumas ponderadas como hacía el modelo general de NTT. Mediante la aplicación de la integral de Choquet para la combinación de los parámetros de calidad el coeficiente de correlación sube a 0,94 y el error medio se reduce en un 10 %.

En [Seshadrinathan and Bovik, 2010] se describe el modelo MOVIE (MOtion-based Video Integrity Evaluation), el cual analiza la degradación en una secuencia de vídeo desde un punto de vista espacial, temporal y espacio-temporal, evaluando la calidad del movimiento a lo largo de distintas trayectorias y distintas escalas. Más concretamente, este algoritmo genera dos componentes de calidad, una asociada a la calidad espacial y otra a la calidad temporal. En primer lugar, las señales de vídeo degradada y de referencia se descomponen mediante un filtro Gabor. La métrica de calidad espacial se basa en una técnica similar a la utilizada en el modelo SSIM [Wang et al., 2004]. La métrica de calidad temporal se genera utilizando la información de movimiento de la señal de referencia. Por último, ambas métricas se combinan para obtener el índice MOVIE.

En [Ou et al., 2011b] se propone un modelo de calidad de vídeo centrado en aplicaciones móviles, considerando resoluciones WVGA (854x480). Mediante este modelo los autores evalúan el efecto de la resolución espacial, la resolución temporal y el nivel de cuantización, concluyendo que la degradación de la calidad con respecto a la resolución temporal es independiente del nivel de cuantización, mientras que la degradación de la calidad con respecto a la resolución espacial es independiente de la resolución temporal y dependiente con respecto al nivel de cuantización.

Una nueva versión de este modelo se puede encontrar en [Ou et al., 2011a] y en [Ma et al., 2012] donde los autores evaluan la dependencia de la calidad en contenidos de resolución CIF con respecto a la resolución temporal y el nivel de cuantización, teniendo en cuenta también el tipo de contenido de cada secuencia. De manera análoga al trabajo anterior, los autores proponen como modelo de calidad el producto de dos funciones, cada una de ellas encargada de modelar el efecto (independiente) de la reso-

lución temporal y el nivel de cuantización, incluyendo como novedad en este artículo, un factor de ajuste que depende del contenido de la secuencia. Resultan de especial interés algunas de las medidas que han utilizado para extraer información sobre el contenido de la secuencia:

- Diferencias entre tramas (absoluta, normalizada y desplazada)
- Magnitud de los vectores de movimiento (MVM)
- Intensidad de la actividad del movimiento (desviación típica de la magnitud de los vectores de movimiento)

En [Wolf and Pinson, 2011] se presenta una evolución del modelo general VQM de NTIA, el denominado modelo VQM_VFD. Hay dos novedades importantes en VQM_VFD con respecto a su predecesor. La primera de ellas consiste en un proceso de calibración mejorado, orientado a detectar variaciones en el retardo entre las tramas de la secuencia original y degradada. La segunda se basa en un entrenamiento del modelo más exhaustivo que en el caso del modelo general, utilizando un amplio rango de resoluciones, que van desde QCIF a resoluciones HD, y un gran número de secuencias de entrenamiento.

VQM_VFD sigue un proceso análogo al del modelo general. Comienza con un proceso de calibración para eliminar ganancias en los valores de luminancia y crominancia y desplazamientos o escalados espaciales (si los hubiera). A este proceso, VQM_VFD incorpora un módulo para contabilizar el retardo relativo (VFD) entre las tramas de las secuencias original y degradada. Con la información VFD el modelo genera una secuencia de referencia "VFD-armonizada", en la cual el efecto del retardo ha sido eliminado. Por ejemplo, si en la señal degradada se pierden o se repiten determinadas tramas, en la señal "VFD-armonizada" se eliminan o se repiten las tramas correspondientes de la señal de referencia.

Con la información del retardo variable, los datos de calibración y la señal "VFD-armonizada" se extraen un conjunto de parámetros de calidad. Estos parámetros incluyen los del modelo general y dos nuevos parámetros:

- VFD_Par1: este parámetro cuantifica los saltos de trama anormales con respecto a la progresión normal de tramas de vídeo a lo largo del tiempo.
- VFD_Par1·PSNR_VFD: el cálculo del PNSR entre la señal original y degradada, puede dar lugar a una penalización excesiva como consecuencia del retardo variable de las tramas entre ambas secuencias. El cálculo del PSNR utilizando la señal "VFD-armonizada" elimina este problema (PSNR_VFD). Sin embargo, el resultado de PSNR_VFD no impone penalizaciones debido a la diferencia de retardo. Teniendo esto en cuenta, el modelo utiliza el producto entre VFD_Par1

y VFD_PSNR para capturar el efecto perceptual conjunto del PSNR y el retardo variable.

Por último, una red neuronal se encarga de generar la predicción de calidad en base a todos los parámetros de calidad extraídos.

Según los autores, el modelo VQM_VFD alcanza un coeficiente de correlación de Pearson mayor de 0,9 para todas las resoluciones contempladas (QCIF, CIF, VGA, SD y HD). En [Besson et al., 2013] se evalúa también el rendimiento de VQM_VFD en comparación con otras métricas de calidad (PSNR, SSIM, modelo general VQM, métricas específicas de SVC y varias métricas NR) aplicadas a secuencias de vídeo codificadas con SVC. De estas métricas, VQM_VFD fue la que mejor resultado consiguió, obteniendo una correlación en torno a 0,81, pese a no estar entrenada específicamente para este tipo de esquema de codificación. En [Wulf and Zolzer, 2013] también se pone de manifiesto las ventajas que supone el nuevo método de calibración VFD en comparación con otras métricas de calidad de vídeo, aplicadas a varias bases de datos de secuencias de vídeo.

4.2.4. Conclusiones extraídas del estado del arte

La principal conclusión que se puede extraer del análisis del estado del arte es que para realizar una monitorización en tiempo real de la calidad percibida en servicios de streaming de vídeo adaptativo OTT, es necesario llevar a cabo el desarrollo de un nuevo modelo sin referencia adecuado a las características de dicho servicio.

Los modelos analizados en el estado del arte no son directamente aplicables por los siguientes motivos:

- La mayoría de ellos están entrenados con contenidos de baja resolución
    - Algunos modelos se centran sobre todo en escenarios móviles, de ahí las resoluciones elegidas. Aun así, en escenarios móviles cada vez es más común que dispositivos como tablets e incluso algunos smartphones, soporten resoluciones Full-HD.
    - En ciertos modelos propuestos en la literatura la dependencia de la calidad con respecto a la tasa de bit de codificación no corresponde con las medidas realizadas en esta tesis. Esto puede deberse también a la diferencia de resoluciones consideradas
- No hay modelos sin referencia estandarizados para resoluciones HD, ni tampoco se han encontrado modelos NR que emulen a algún modelo de referencia completo.
- Algunos de los modelos NR HD están entrenados con muy pocas secuencias de

vídeo (menos de 10 en algunos casos), por lo que su validez y su aplicabilidad son limitadas.

4.3. Desarrollo del modelo

En esta sección se describe cómo se ha abordado el diseño y el desarrollo de la parte del modelo que estima la degradación en la calidad percibida introducida por la fase de codificación.

A la hora de desarrollar un modelo de calidad de vídeo sin referencia se pueden plantear dos estrategias distintas para alcanzar el objetivo final del modelo: obtener una predicción de la calidad percibida por los usuarios.

La primera estrategia consistiría en seleccionar y/o generar un conjunto de secuencias de vídeo de entrenamiento, diseñar y llevar a cabo tests de evaluación subjetiva donde un conjunto de usuarios valorarían la calidad de estas secuencias de vídeo y por último desarrollar un algoritmo que permitiese predecir las valoraciones obtenidas en los tests subjetivos utilizando solo la información de la señal de vídeo degradada.

Además de la dificultad del diseño del algoritmo, no es despreciable el esfuerzo y los recursos asociados a la realización de los tests subjetivos necesarios para obtener las valoraciones de calidad que el algoritmo debe predecir.

La segunda estrategia, utilizada en esta tesis, consiste en variar ligeramente el objetivo del modelo. En la primera estrategia el objetivo del modelo es predecir las valoraciones de los usuarios, mientras que lo que se propone en la segunda estrategia es predecir el resultado que generaría un modelo de referencia completa, resultado que se supone es una buena predicción de las valoraciones de los usuarios. Así pues, se sustituye la fase de evaluación subjetiva por una fase en la que se "mide" la calidad percibida mediante un modelo de referencia completa.

Así pues, se han llevado a cabo las siguientes tareas:

1. Elección de un modelo de calidad de vídeo de referencia completa o referencia reducida: este modelo será tomado como el modelo "objetivo" de nuestro modelo sin referencia. Expresado de otro modo, el modelo que se desarrolla en esta tesis deberá ofrecer resultados similares a los que ofrezca el modelo de referencia completa/reducida seleccionado, pero utilizando como input únicamente el vídeo degradado.

2. Elección de una base de datos de secuencias de vídeo de prueba: es fundamental contar con un conjunto de secuencias de vídeo, lo suficientemente amplio y variado, que permita llevar a cabo el entrenamiento del modelo.

3. Medidas de la calidad percibida de los vídeos de la base de datos de prueba

codificados a distintas tasas de bit y ajuste a una función matemática de la forma $Q = f(bitrate)$ y extracción de parámetros de la curva.

4. Entrenamiento del modelo: desarrollo de un mecanismo que permita obtener los parámetros de la curva $Q = f(bitrate)$ en función del vídeo degradado (sin referencia).

5. Evaluación del modelo.

4.3.1. Selección del modelo de referencia

Como se ha comentado en la sección anterior, en el desarrollo de este modelo de calidad de vídeo se han sustituido las valoraciones subjetivas de calidad por el resultado generado por un modelo de referencia completa. A continuación se describe el modelo seleccionado y las razones que han llevado a tal decisión.

Tras analizar el estado del arte, **se ha decidido utilizar como modelo de referencia el modelo VQM_VFD**. El modelo VQM_VFD [Wolf and Pinson, 2011] es una evolución del modelo general VQM desarrollado por NTIA, el cual está adaptado a un amplio rango de resoluciones, nuevos tipos de degradaciones y ha sido entrenado utilizando un amplio conjunto de secuencias de vídeo. Como se vio en el estado del arte, el rendimiento de este modelo está contrastado por diversos estudios de la literatura, lo cual lo valida para ser utilizado como modelo objetivo en esta tesis.

Otro aspecto importante, de carácter práctico, que hace que VQM_VFD se posicione por delante de otros modelo que podrían haber sido utilizados (como el modelo VQuadHD [ITU, 2011b]) es su carácter abierto y la disponibilidad de implementaciones.

4.3.2. Selección de la base de datos de secuencias de vídeo de prueba

La tabla 4.2 recoge algunas de las bases de datos de secuencias de vídeo de prueba más destacadas de las que se han podido encontrar en la literatura.

Para el desarrollo de esta tesis se ha seleccionado la base de datos VQEG HDTV [VQEG, 2011], ya que incluye la mayor parte de las secuencias de vídeo de prueba utilizadas en los experimentos del VQEG en los que se evaluaron los últimos modelos de calidad estandarizados por ITU. Esto garantiza la validez y la adecuación de esta base de datos de secuencias de vídeo a esta tesis doctoral. A continuación se ofrecen algunos detalles del diseño de dicha base de datos.

La base de datos de vídeos está formada por 5 colecciones de vídeos con formato 1080p y 1080i a 25 y 30 fps y 10 segundos de duración cada una. El contenido de dichas secuencias es representativo de un amplio conjunto de aplicaciones:

- Películas y trailers de películas

Tabla 4.2: Bases de datos de secuencias de vídeo de prueba HD

Nombre	Número de vídeos	Resolución	Formato	Evaluaciones subjetivas	Catálogo de vídeos
LIVE Video Quality Database	150	768x432	YUV 4:2:0 y versiones comprimidas en H.264 y MPEG-2	Sí	Vídeo original, distorsión wireless, distorsión IP, compresión H.264 y compresión MPEG-2
The Consumer Digital Video Library	2500	De CIF a 1080p	Contenedores AVI con YUV	No	Agrupa varios catálogos de vídeos
IRCCyN/IVC 1080i	192	1080i	YUV 4:2:2	Sí	Vídeo original y compresiones en H.264
VQEG HDTV	30	1080p y 1080i	Contenedor AVI YUV 4:2:2	Sí	Vídeo original y diversos HRC (con pérdidas y sin pérdidas)
JEG264HMIX1 database	170	1080p	Contenedor AVI YUV 4:2:2	Sí	Vídeo original y diversos HRC (con pérdidas y sin pérdidas)
IVP Subjective Quality Video Database	180	1080p	YUV 4:2:0	Sí	Vídeo original, H.264, MPEG2, Dirac coding, IP error

- Deportes
- Vídeos musicales
- Anuncios publicitarios
- Películas de animación (dibujos animados)
- Noticiarios
- Vídeos caseros
- Material general de televisión (documentales, series, etc.)

Para cada colección de vídeos se cumplen las siguientes características:

- Todos los vídeos tienen la misma tasa de frames.
- Todos los vídeos son progresivos o todos son entrelazados.
- Al menos uno de los vídeos debe ser muy fácil de codificar.
- Al menos uno de los vídeos debe ser muy difícil de codificar.
- Al menos uno de los vídeos debe incluir muchos detalles espaciales.
- Al menos uno de los vídeos debe contener mucho movimiento o cambios rápido de escenas.
- Si es posible, uno de los vídeo debe tener múltiples objetos moviéndose de manera aleatoria e impredecible.
- Al menos uno de los vídeos debe ser muy colorido.
- Si es posible uno de los vídeos contendrá alguna animación superpuesta (texto que se desplaza, por ejemplo).
- Si es posible, al menos un vídeo tendrá poco contraste.
- Si es posible, al menos un vídeo tendrá mucho contraste.
- Si es posible, al menos un vídeo tendrá poco brillo.
- Si es posible, al menos un vídeo tendrá mucho brillo.

4.3.3. Medidas de VQM-VFD

Una vez seleccionada la base de datos de secuencias de vídeo, la siguiente fase es medir VQM_VFD en cada uno de las secuencias de vídeo de entrenamiento, utilizando el software que proporciona NTIA [NTIA, 2011].

Más concretamente, para estudiar el efecto que tiene la tasa de bit de codificación, cada uno de los vídeos de la base de datos se ha codificado a distintas tasas de bit de codificación. Las tasas de bit elegidas han sido 1, 1,5, 2, 4, 8 y 12 Mbit/s. Así pues, tras codificar cada vídeo a las tasas indicadas anteriormente, se mide VQM_VFD en cada uno de ellos.

El siguiente paso es intentar obtener una relación entre VQM_VFD y la tasa de bit de codificación. El método seguido en esta tesis ha sido el siguiente:

1. Representar, para cada vídeo, los puntos obtenidos en la fase de medición anterior. Puntos de la forma $(BitRate_i, VQM\_VFD_i)$.

2. Ajustar los puntos a una curva.

3. Seleccionar el tipo de función que mejor ajuste proporcione.

A continuación se presentan los resultados de las medidas de VQM_VFD.

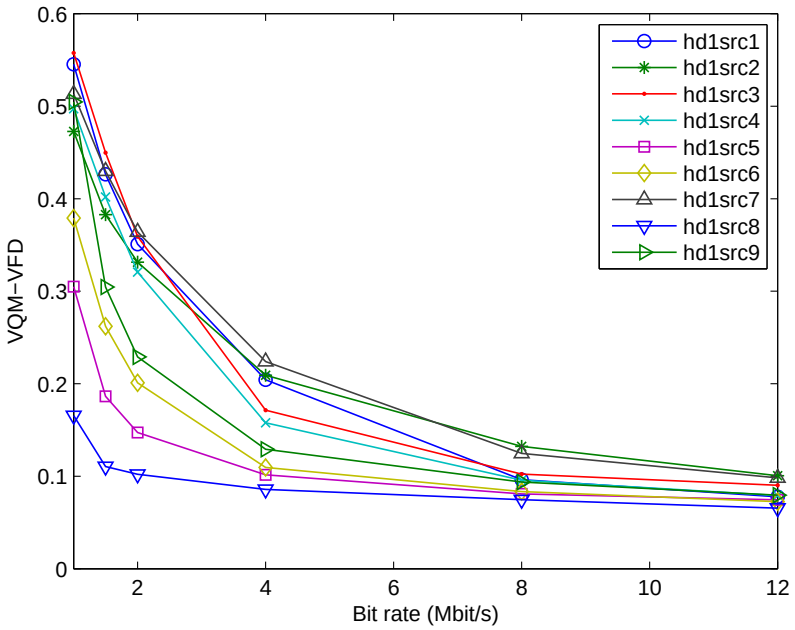


Figura 4.3: VQM_VFD para las secuencias de vídeo VQEG-HD1

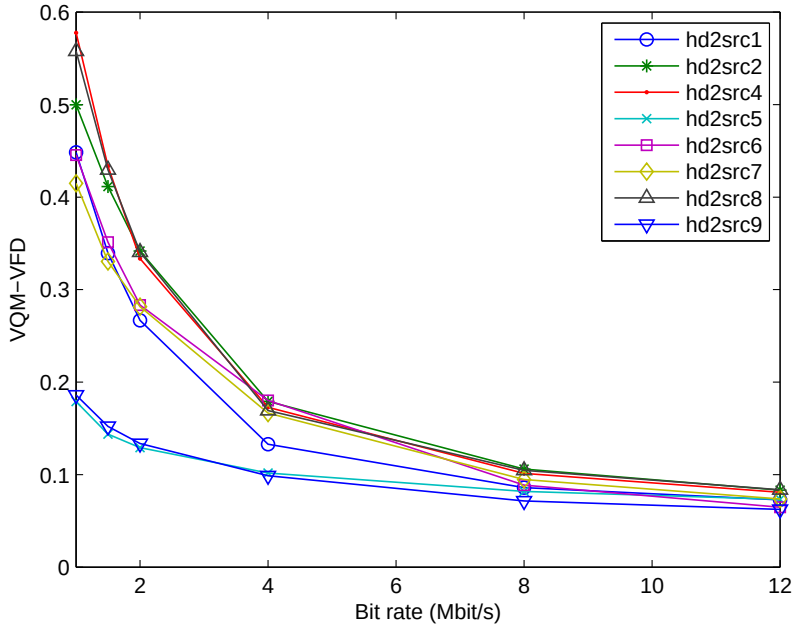


Figura 4.4: VQM_VFD para las secuencias de vídeo VQEG-HD2

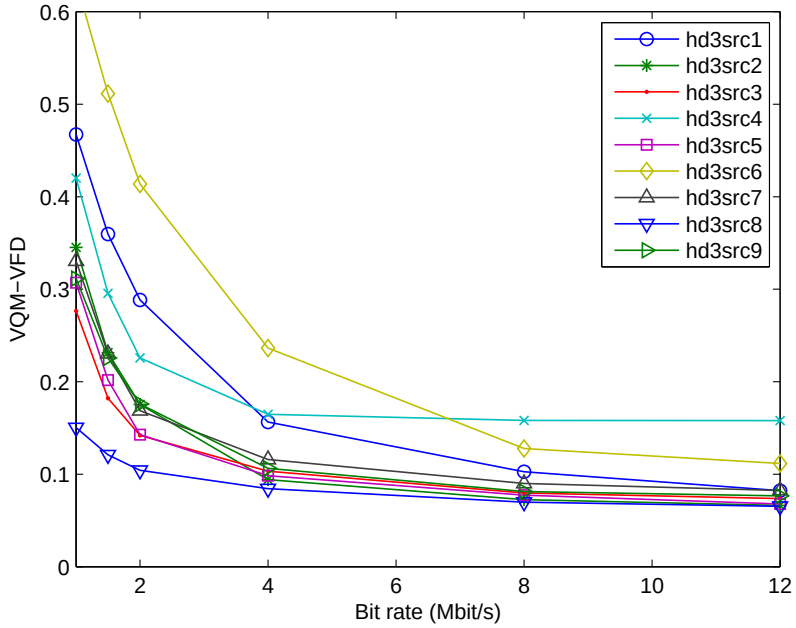


Figura 4.5: VQM_VFD para las secuencias de vídeo VQEG-HD3

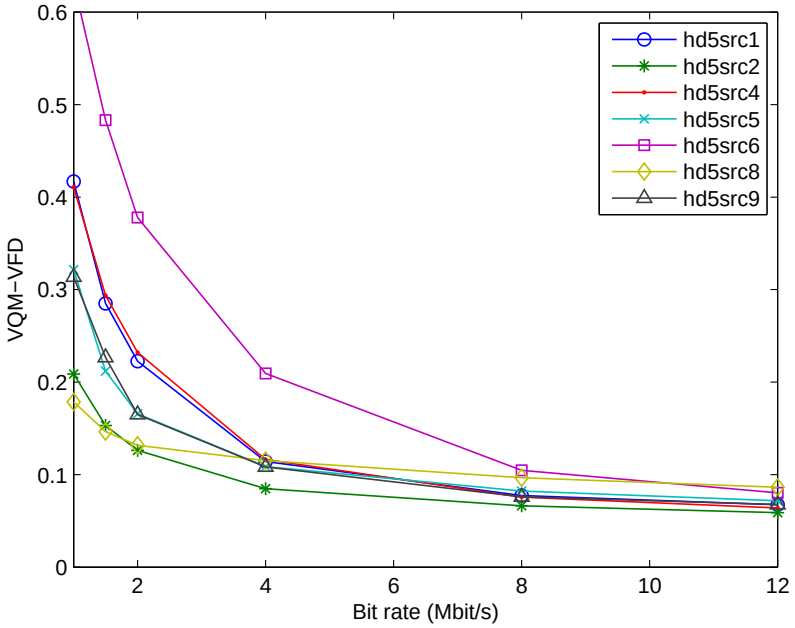


Figura 4.6: VQM_VFD para las secuencias de vídeo VQEG-HD5

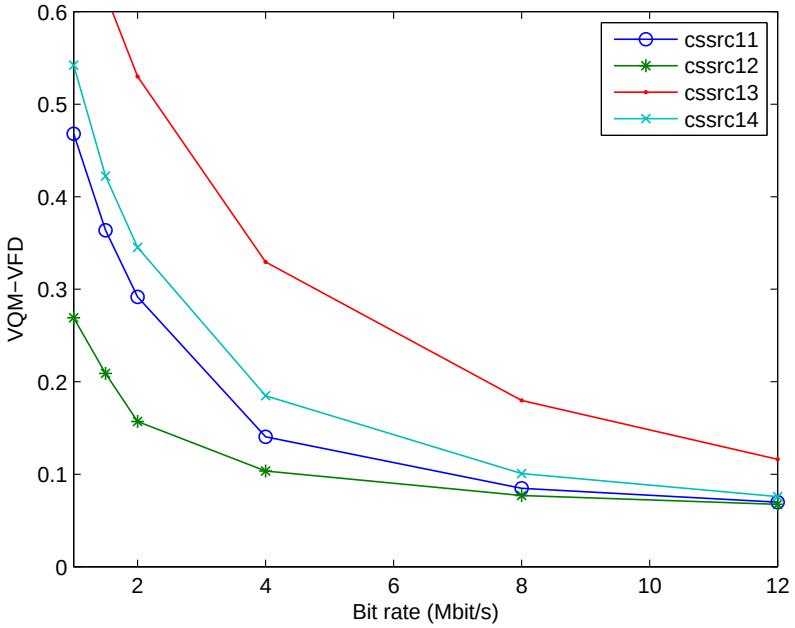


Figura 4.7: VQM_VFD para las secuencias de vídeo VQEG-CommonSet

A la vista de las gráficas obtenidas, y tras evaluar diferentes formas de curva, se puede concluir que la variación de VQM_VFD se ajusta correctamente a una función potencial con respecto a la tasa de codificación. Además, este resultado es acorde a los resultados obtenidos [de la Cruz Ramos, 2012] y en [de la Cruz Ramos et al., 2012] para el caso del modelo general VQM de NTIA.

Así pues, la expresión general de VQM_VFD en función de la tasa de bit de codificación que se propone es la siguiente:

$$VQM\_VFD = a \cdot bitRate^{b} \tag{4.12}$$

En la ecuación 4.12, a y b son parámetros de ajuste. En las siguientes tablas se incluyen los parámetros a y b que se han obtenido al realizar el ajuste para cada una de las secuencias de vídeo analizadas. Además se incluye el valor del coeficiente de determinación (R^2), que evalúa la bondad del ajuste de la curva a los puntos medidos.

Tabla 4.3: Parámetros de ajuste VQM_VFD para las secuencias VQEGHD

Secuencia	a	b	R^2
hd1src1	0,58647226	-0,8211043	0,99197349
hd1src2	0,49254306	-0,63136378	0,99780762
hd1src3	0,58009557	-0,79239125	0,98531872
hd1src4	0,51981897	-0,78772016	0,99044087
hd1src5	0,24568266	-0,53341738	0,92598587
hd1src6	0,33822496	-0,67056532	0,96777997
hd1src7	0,55530542	-0,69373693	0,99317233
hd1src8	0,13980649	-0,31994965	0,90250549
hd1src9	0,4210787	-0,72370215	0,96358547
hd2src1	0,4426828	-0,76685007	0,98787554
hd2src2	0,53556758	-0,76012725	0,99397998
hd2src4	0,58013391	-0,81926093	0,99578007
hd2src5	0,16965816	-0,34951663	0,989855
hd2src6	0,47852074	-0,78986601	0,99235916
hd2src7	0,4380379	-0,71762124	0,99661811
hd2src8	0,56731323	-0,79730321	0,99334733
hd2src9	0,18312982	-0,44235467	0,9986254
hd3src1	0,46600476	-0,7197805	0,99457392
hd3src2	0,29968754	-0,67195263	0,95475447
hd3src3	0,23169026	-0,50687599	0,94204284

continúa en la siguiente página

Tabla 4.3 – continuación

Secuencia	a	b	R^2
hd3src4	0,3429309	-0,37525777	0,83187022
hd3src5	0,25466332	-0,58017774	0,94046734
hd3src6	0,66439621	-0,74439482	0,99252885
hd3src7	0,28363219	-0,54655159	0,94842379
hd3src8	0,13928317	-0,3266249	0,97074915
hd3src9	0,27952086	-0,57671319	0,96085921
hd5src1	0,38245344	-0,75047743	0,98023699
hd5src2	0,18957983	-0,50337983	0,97544947
hd5src4	0,39286564	-0,7732192	0,98743862
hd5src5	0,27587664	-0,58271764	0,96386709
hd5src6	0,66547057	-0,85931685	0,99719697
hd5src8	0,16807618	-0,27204159	0,97860912
hd5src9	0,28264325	-0,61784556	0,97509032
cssrc11	0,4811712	-0,80928367	0,99114326
cssrc12	0,25109565	-0,56284161	0,97952443
cssrc13	0,82076902	-0,74245659	0,98177553
cssrc14	0,57436999	-0,81897775	0,99725283

Como se puede observar en las tablas, el ajuste del modelo es muy preciso. El valor medio del coeficiente de determinación de todas las secuencias es de 0,973266617 con una varianza de 0,001068836. Siendo el valor máximo de R^2 igual a 0,998625398 y el valor mínimo igual a 0,831870223.

Este valor mínimo del coeficiente de determinación se obtiene en la secuencia hd3src4 (secuencia 4 de la colección HD3), para la cual, como se puede ver en la figura 4.8, VQM_VFD obtiene valores peculiares que no siguen la misma tendencia que el resto de secuencias.

Como se puede ver en la gráfica, el valor de VQM_VFD para esta secuencia se satura con el aumento de la tasa de codificación. Esto es debido a que a bajas tasas de codificación, dada la naturaleza de la secuencia (fragmento de una serie de dibujos animados), el codificador obtiene ya niveles de calidad bastante aceptables. Al aumentar la tasa de codificación, la ganancia en cuanto a calidad es casi imperceptible, solo mejorando en aquellos fotogramas de vídeo con mayor nivel de movimiento, por lo que los valores de VQM_VFD apenas varían.

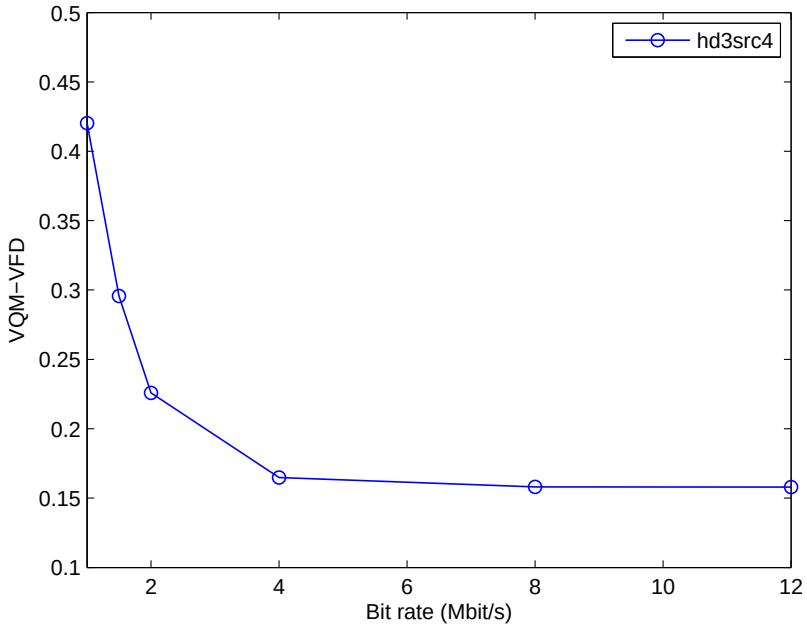


Figura 4.8: VQM_VFD para la secuencia de vídeo VQEG-HD3SRC4

4.3.4. Entrenamiento del modelo

4.3.4.1. Consideraciones iniciales

Una vez fijada la expresión analítica del modelo, el objetivo del siguiente paso es plantear un método que permita, a partir de un vídeo codificado a una tasa de bit determinada, predecir el valor que tendrán los parámetros a y b del modelo, obteniendo a partir de ellos una estimación de VQM_VFD según la ecuación 4.12, y por tanto una estimación de la calidad percibida.

El enfoque propuesto en esta tesis para el desarrollo de esta fase del modelo se basa en utilizar como variables de predicción diversas magnitudes relacionadas con el contenido y la complejidad del contenido de la secuencia de vídeo degradada. Este enfoque ha sido aplicado en diversos trabajos de la literatura, entre los que destacan [de la Cruz Ramos, 2012], [Ou et al., 2011a] y [Ma et al., 2012].

Además de seleccionar las variables de predicción que se van a utilizar, es necesario encontrar un método o una técnica que permita procesar y combinar estas variables de predicción para obtener un valor aproximado de los parámetros a y b.

En el resto de esta subsección se describen las distintas estrategias que han sido evaluadas (y algunas descartadas) para generar las predicciones de a y b.

La primera estrategia analizada consiste en aplicar un enfoque similar al propuesto en [de la Cruz Ramos, 2012], donde se plantea un método de estimación de parámetros

basado en interpolación sobre una tabla con valores precalculados de las variables de estimación ASI y ATI. Sin embargo, este mecanismo no se ha podido aplicar directamente en esta tesis, ya que para ciertas secuencias de entrenamiento con valores similares de ASI y ATI se obtienen resultados de los parámetros a y b muy diferentes. Además, las secuencias que mostraban este comportamiento no eran directamente clasificables en grupos que permitiesen armonizar comportamientos tan dispares (como se propone en la tesis de P. de la Cruz). Para ilustrar este problema, a continuación se presentan dos gráficas en las que se muestran los valores de a y b en función de ASI y ATI.

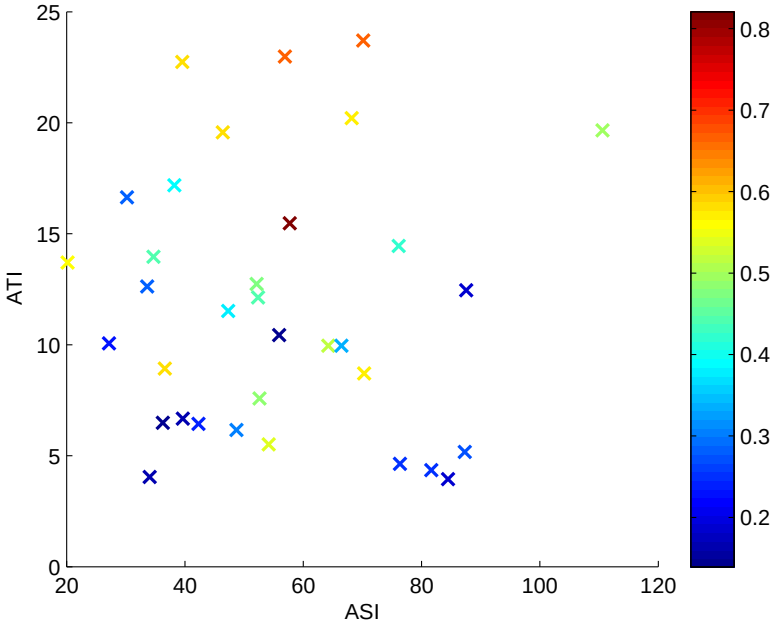


Figura 4.9: Valores del parámetro a en función de ASI y ATI

Como puede verse en las figuras 4.9 y 4.10, puntos (secuencias de vídeo) muy cercanos en el plano ASIxATI tienen valores de a y b muy distintos entre sí. El ejemplo más claro son los puntos $S_1 = (33, 62, 12, 62)$ y $S_2 = (34, 67, 13, 96)$. Para el caso del parámetro a, los valores de a asociados a S_1 y S_2 son 0,28 y 0,44 respectivamente. Para el caso del parámetro b, los valores de b asociados a S_1 y S_2 son -0,58 y -0,72 respectivamente. Así pues, mientras que la distancia entre los valores de ASI y ATI asociados a S_1 y S_2 es de menos del 1 % del rango de cada variable, la distancia entre los valores de a y b es de aproximadamente el 24 % de su rango.

Este resultado invalida la posibilidad de utilizar la técnica de la interpolación para predecir los valores de a y b en función de ASI y ATI.

Además del método de interpolación se han analizado otros métodos de ajuste como

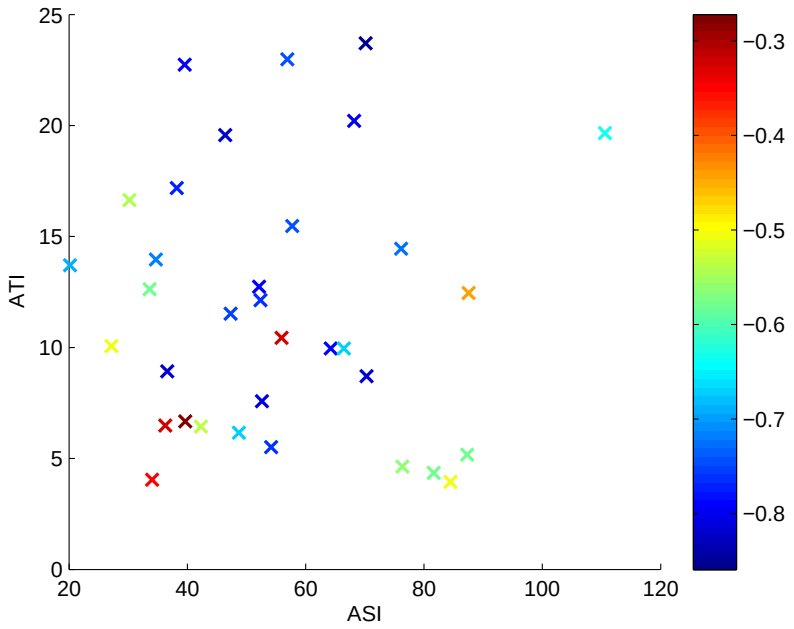


Figura 4.10: Valores del parámetro b en función de ASI y ATI

la regresión lineal múltiple (con términos de interacción y términos cuadráticos), como el utilizado en el modelo de [Ou et al., 2011a]. Sin embargo, estos métodos tampoco han proporcionado resultados adecuados.

4.3.4.2. Enfoque basado en aprendizaje automático

La pobreza de los resultados obtenidos utilizando los métodos anteriormente descritos pone de manifiesto que la relación entre a y b y los valores de ASI y ATI puede ser no lineal y también sugiere la posibilidad de que se precisen más variables de predicción para obtener una estimación de a y b con un error asumible.

Tras analizar diversas opciones, el modelo propuesto se basa en la utilización de una red neuronal cuyas variables de entrada son un conjunto de características del contenido y de la complejidad del contenido de la secuencia de vídeo codificada. Las redes neuronales permiten, con la arquitectura y el entrenamiento adecuados, modelar relaciones muy complejas y no lineales entre las variables de entrada y de salida. Esta característica es especialmente interesante para los objetivos de esta tesis, ya que como se ha comentado anteriormente, la relación existente entre las variables de predicción (características de la secuencia de vídeo) y los parámetros del modelo que se desean estimar (parámetros a y b de la curva $VQM\_VFD = a \cdot bitRate^b$) no siguen una relación lineal sencilla.

Las redes neuronales son una herramienta bien conocida en el ámbito del aprendizaje automático y se utilizan principalmente para el reconocimiento automático de patrones. Una red neuronal se suele presentar como un sistema de neuronas interconectadas que calculan un resultado a partir de unos datos de entrada mediante la colaboración de las neuronas que forman la red (figura 4.11).

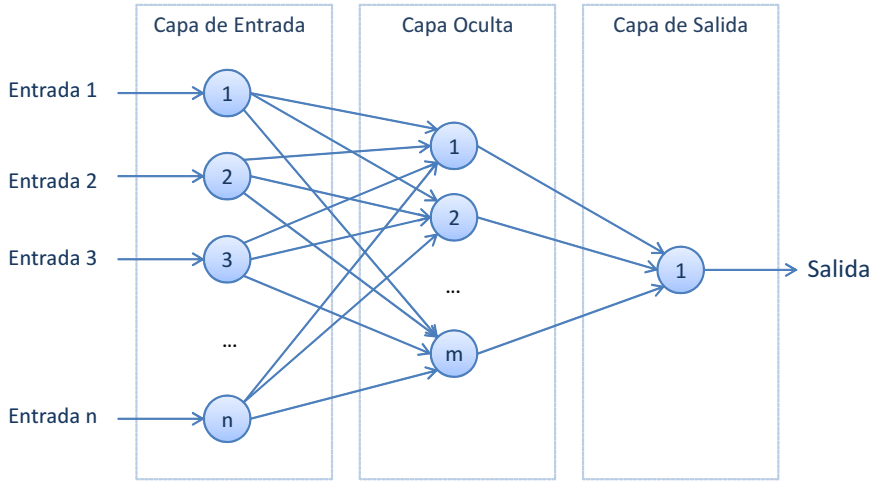


Figura 4.11: Red neuronal: perceptrón multicapa

Las redes neuronales se pueden definir mediante los siguientes parámetros:

- El patrón de interconexión entre las distintas capas de neuronas.
- El proceso de aprendizaje utilizado para actualizar el peso de las interconexiones entre neuronas.
- La función de activación que permite a la neurona generar un resultado a partir de las variables de entrada.

En el diseño de una red neuronal hay varios aspectos clave que se deben tener en cuenta:

- Arquitectura de la red neuronal
- Selección de las variables de entrada
- Preprocesado de los datos de entrenamiento

La arquitectura que se ha utilizado en esta tesis corresponde a un perceptrón multicapa. Esta arquitectura está formada por varias capas de neuronas interconectadas una tras otra, formando un grafo dirigido desde la capa de entrada a la capa de salida. Excepto los nodos de entrada, todos los nodos del grafo son neuronas con una

función de activación no lineal. Este tipo de red se entrena utilizando algoritmos de retropropagación o propagación hacia atrás, como el algoritmo Levenberg-Marquardt, regularización bayesiana o el método del gradiente conjugado escalado.

La arquitectura de la red neuronal utilizada en esta tesis se puede ver en la figura 4.12.

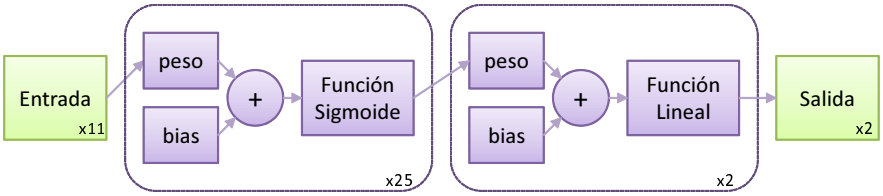


Figura 4.12: Arquitectura de la red neuronal utilizada en el modelo

La red neuronal propuesta consta de una primera capa formada por los nodos de entrada, encargados de recibir las variables de entrada, las cuales se describirán más adelante. Cada nodo de la capa de entrada está conectado a todos los nodos de la capa oculta, la cual consta de 25 neuronas, cuya función de activación es una función sigmoide:

$$g(x) = \frac{1}{1 + e^{-x}} \tag{4.13}$$

El resultado producido por las neuronas de la capa oculta se utiliza como entrada para 2 neuronas (con función de activación lineal) en la capa de salida, que son las encargadas de proporcionar la estimación final de los parámetros a y b del modelo. Además, como suele ser habitual en el diseño de redes neuronales, se incluye un nodo de ajuste (o bias, como se suele denominar en la literatura) tanto en la capa oculta como en la capa de salida cuya utilidad principal es la de mejorar el proceso de aprendizaje, introduciendo un grado de libertad extra en cada capa.

En cuanto al número de neuronas utilizado en cada capa, la decisión adoptada está relacionada con uno de los problemas más habituales en el ámbito del aprendizaje automático: la búsqueda de un compromiso entre el error y el sobre ajuste, o como se conoce en la literatura "bias-variance dilemma" o "bias-variance tradeoff" [Geman et al., 1992].

En general, las técnicas como la regresión lineal o las redes neuronales deben alcanzar un equilibrio entre el error obtenido en el proceso de aprendizaje y la capacidad para predecir adecuadamente nuevos valores a partir de datos de entrada no utilizados en el entrenamiento. Las redes neuronales son capaces, con la arquitectura adecuada y con el número suficiente de neuronas, de "memorizar" cada uno de los datos de entrenamiento, obteniendo así un error nulo en el entrenamiento. Sin embargo, esto puede

dar lugar a que la función matemática que está representando la red neuronal sea tan compleja, que aún ajustándose perfectamente a los datos de entrenamiento, presente un error muy elevado cuando a la red neuronal se le presentan nuevos inputs, es decir, la red neuronal ha "memorizado" los datos de entrenamiento pero no es capaz de generalizar los resultados cuando se utilizan datos de entrada distintos a los utilizados en el entrenamiento.

Este comportamiento se puede ver más claramente con el ejemplo simplificado que se muestra en la figura 4.13

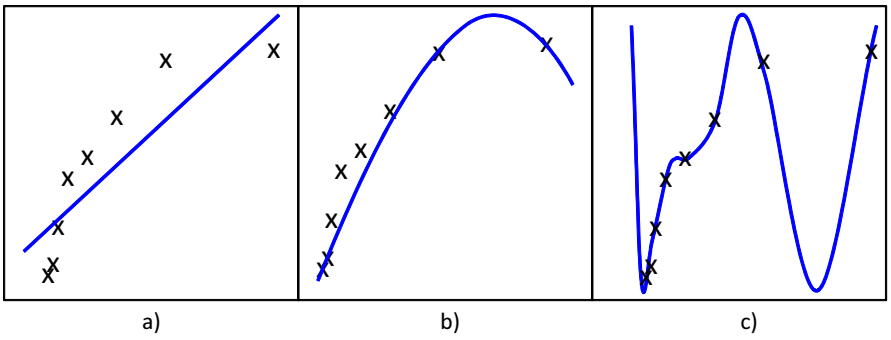


Figura 4.13: Bias y overfitting

En este ejemplo, se está desarrollando un modelo de predicción simplificado para estimar la variable del eje y a partir de la del eje x. Para ello, se utilizan una serie de datos de entrenamiento, con los que se va a entrenar algún algoritmo de aprendizaje automático (regresión lineal, red neuronal, etc.). A continuación se analiza cada uno de los casos representados en la figura:

- En el primer caso (subfigura a), el modelo que se ha diseñado es demasiado simple. El algoritmo ha generado una función lineal que obtiene un alto error con respecto a los datos de entrenamiento y probablemente también con respecto a nuevos datos. En este caso se dice que el modelo sufre una cierta desviación (bias).

- En el tercer caso (subfigura c), se puede ver que el error que comete el modelo con respecto a los datos de entrenamiento es nulo. La función matemática que ha generado el algoritmo es tan compleja que todos los puntos de entrenamiento coinciden exactamente con la función. Sin embargo, a la vista de la tendencia general que siguen los datos, no parece que este modelo vaya a ofrecer buenos resultados cuando sea alimentado con nuevos datos diferentes a los usados en el entrenamiento. En este caso se dice que el modelo está sobre-ajustado o sobre-entrenado, es decir, el modelo sufre overfitting.

- Por último, en el caso de la subfigura b, se ha encontrado un compromiso entre el error cometido en la fase de entrenamiento (hay algunos puntos que no están exactamente sobre la curva) y el nivel de ajuste del modelo, permitiendo así que los resultados obtenidos con nuevos inputs sean válidos.

Para evaluar y controlar el efecto del overfitting se suele utilizar una técnica que consiste en separar los datos de entrenamiento disponibles en tres grupos:

- Conjunto de entrenamiento
- Conjunto de validación
- Conjunto de prueba

Una vez que se han separado los datos, se inicia el algoritmo de entrenamiento (utilizando únicamente el conjunto de datos de entrenamiento) y se van creando unas gráficas en cada iteración del algoritmo de entrenamiento. Estas gráficas muestran el error con respecto al conjunto de entrenamiento y también el error con respecto al conjunto de validación (que no se ha utilizado en el entrenamiento). A estas gráficas se les conoce como curvas de aprendizaje (learning curves). Un ejemplo de curva de aprendizaje se puede ver en la figura 4.14.

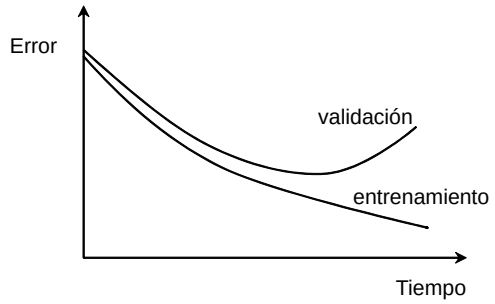


Figura 4.14: Curva de aprendizaje

Como se puede ver en la figura anterior, al iniciar el entrenamiento tanto el error con respecto al conjunto de datos entrenamiento como el error con respecto al conjunto de datos de validación (no utilizados en el algoritmo de entrenamiento) es alto. Al progresar el entrenamiento, ambos errores van disminuyendo. Sin embargo, se alcanza un punto en el proceso de entrenamiento en el que el error con respecto al conjunto de datos de validación deja de decrecer y empieza a ser cada vez mayor. Este punto de inflexión indica que se está empezando a producir overfitting. A partir de este punto, si se sigue con el entrenamiento, el modelo será cada vez más complejo (adaptándose mejor a los datos de entrenamiento) e irá perdiendo la capacidad de generalizar resultados para

datos no utilizados en el entrenamiento. Los datos del conjunto de prueba se utilizan para evaluar el resultado final del entrenamiento.

Por último, dependiendo de si el modelo en cuestión sufre de bias o de overfitting se deben aplicar una serie de medidas u otras. La tabla 4.4 es un resumen con algunas técnicas a aplicar en casa caso.

Tabla 4.4: Técnicas de reducción de bias y overfitting

Reducir bias	**Reducir overfitting**
Aumentar el número de variables de entrada	Utilizar más datos de entrenamiento
Aumentar la complejidad del modelo (más neuronas en la red neuronal, polinomios de mayor grado en regresión lineal, etc.).	Reducir el número de variables de entrada
Reducir el parámetro de regularización	Aumentar el parámetro de regularización

Aprovechando los conceptos relacionados con el aprendizaje automático que se han introducido hasta el momento, se puede justificar la decisión de optar por una técnica de entrenamiento diferente a las analizadas anteriormente en [de la Cruz Ramos, 2012], [Ou et al., 2011a] y [Ma et al., 2012]. Así pues, la utilización de una red neuronal y la utilización de un mayor número de variables de predicción tienen como objetivo disminuir el error de las predicciones, es decir, reducir bias.

4.3.4.3. Variables de predicción

Una vez introducidos estos conceptos relacionados con el aprendizaje automático y justificada la arquitectura de red neuronal elegida, el siguiente paso es analizar las variables de entrada que se van a utilizar como variables de predicción.

La lista de variables utilizadas es la siguiente:

- Información espacial, SI
- Información temporal, TI
- Información espacial media, ASI
- Información temporal media, ATI
- Entropía media, H_{avg}
- Entropía máxima, H_{max}
- Información temporal media de bordes, ATI-Sobel

- Variación sobre la información temporal media de bordes, ATI-Sobel-2
- Módulo medio de los vectores de movimiento, μ_{MVM}
- Coherencia del movimiento, σ_{DVM}
- Cociente entre el módulo medio y la coherencia del movimiento μ_{MVM}/σ_{DVM}

Como se ha comentado, en las primeras fases de desarrollo del modelo se intentó utilizar como input de la red neuronal únicamente las medidas de SI, TI, ASI y ATI. Sin embargo, el error que se obtenía en la fase de entrenamiento de la red neuronal era muy elevado, por lo que se decidió incrementar el número de variables para reducir el bias. A continuación se describe cada una de estas variables con más detalle.

Información espacial y temporal En [Webster et al., 1993] se definen dos medidas perceptuales (es decir, basadas en un modelo perceptual del sistema visual humano) del contenido de información de secuencias de vídeo. Estas medidas fueron posteriormente normalizadas por ANSI en ANSI T1.801.03–1996 [ANSI, 1996] y por ITU en ITU-T P.910 [ITU, 2008f], donde se recomiendan como criterios para la clasificación de secuencias de vídeo en función de su contenido de información.

La Información Espacial SI es una medida de complejidad espacial, es decir, mide la cantidad de detalle espacial percibido por un observador humano en una imagen o secuencia de vídeo. Es usualmente mayor para escenas más complejas espacialmente, y es sensible a cambios en la definición de los bordes de las imágenes, tales como los causados por la borrosidad (blurriness), ruido (noise), teselación (tiling) y distorsión de bloques (block distortion). Se define como:

$$SI = max_n\{std_s\{Sobel(F_n)\}\} \tag{4.14}$$

En la ecuación 4.14 $Sobel(F_n)$ es el resultado de aplicar el filtro de Sobel a los valores de luminancia de la trama F_n, std_s es la desviación típica de los valores de luminancia de la trama filtrada y max_n es el valor máximo de la serie temporal.

Como se puede ver, la definición de SI está basada en el filtrado Sobel. El filtro de Sobel [Jain, 1989] es un sencillo filtro paso-alto de 3x3 píxeles, ampliamente utilizado para la detección de bordes en el ámbito del procesamiento de imágenes.

La Información Temporal TI es una medida de complejidad temporal, es decir, mide la cantidad de cambio temporal percibido por un observador humano en una secuencia de vídeo. Normalmente TI es mayor para escenas con mucho movimiento, y es sensible a las degradaciones en el flujo de movimiento, tales como las causadas por el ruido (noise) y por la pérdida o repetición de tramas. Se define según la ecuación 4.15, donde

ΔF_n es la diferencia píxel a píxel de los valores de luminancia de las tramas F_n y F_{n-1}.

$$TI = max_n\{std_s\{\Delta F_n\}\} \tag{4.15}$$

Es importante destacar que ninguna de estas dos medidas pretende ser una medida de la entropía de la imagen o escena, ni están relacionadas con el contenido de información en el sentido de la teoría de la comunicación, sino que intentan medir la cantidad de detalle espacial y temporal percibidos por un observador humano.

Información espacial y temporal media En [de la Cruz Ramos et al., 2012] se define una variación de las medidas SI y TI con el objetivo de suavizar el efecto que algunas tramas con valores extremos (distorsiones de corta duración, cambios de escena, o tramas erróneas) pueden tener sobre el resultado de dichas medidas. Así pues, se definen los valores ASI y ATI mediante las siguientes ecuaciones 4.16 y 4.17, donde avg_n es el promedio de la serie temporal.

$$ASI = avg_n\{std_s\{Sobel(F_n)\}\} \tag{4.16}$$

$$ATI = avg_n\{std_s\{|\Delta F_n|\}\} \tag{4.17}$$

Medidas de entropía Se han introducido en el modelo dos variables relacionadas con la entropía de la secuencia de vídeo: H_{avg} y H_{max}.

La primera de ellas, H_{avg} es el promedio de la entropía de la componente de luminancia de cada trama de la secuencia de vídeo.

$$H_{avg} = avg_n\{H(F_n)\} \tag{4.18}$$

La segunda, H_{max} es el valor máximo de la entropía de todas las tramas del vídeo.

$$H_{max} = max_n\{H(F_n)\} \tag{4.19}$$

La entropía de cada trama es una medida estadística de la aleatoriedad de la imagen y es una medida que sirve para caracterizar la textura de la imagen [Pham, 2012]. Esta es la principal razón por la que se incluyen estas dos medidas ya que la entropía de la imagen sirve como indicador de la complejidad de codificación de la misma. Se utiliza tanto un valor medio, representativo de la secuencia de vídeo al completo, como el valor máximo de la secuencia, que sirve para tener en cuenta si existe una trama de vídeo especialmente compleja.

El estudio de las texturas de la imagen en codificación de vídeo es un campo estudiado sobre todo con el objetivo de diferenciar zonas de la imagen que se puedan

codificar de manera más burda, aprovechando las particularidades del sistema visual humano [Ndjiki-Nya et al., 2003].

Información temporal media sobre información de bordes El objetivo de estas medidas es obtener un indicador de la cantidad de movimiento que sufren los bordes de los objetos representados en cada trama. Para ello, en primer lugar se lleva a cabo un filtrado Sobel para cada trama (para detectar los bordes) y posteriormente se realiza la diferencia píxel a píxel entre las tramas filtradas. Las dos medidas difieren en el estadístico utilizado para su computación. En una de ellas se utiliza un promedio y en la otra la desviación típica. Así pues estas variables se han definido de según las ecuaciones 4.20 y 4.21.

$$ATISobel = avg_n\{std_s\{\Delta Sobel(F_n)\}\} \tag{4.20}$$

$$ATISobel2 = avg_n\{avg_s\{\Delta Sobel(F_n)\}\} \tag{4.21}$$

Medidas relacionadas con vectores de movimiento Por último, se utilizan tres medidas relacionadas con los vectores de movimiento que calcula el codificador. A diferencia de las medidas obtenidas mediante diferencias entre tramas contiguas, los vectores de movimiento ofrecen información de trayectorias de objetos que en general involucra varias tramas del vídeo. Las medidas utilizadas en el modelo son:

- Módulo medio μ_{MVM}: promedio del módulo de los vectores de movimiento.
- Coherencia de movimiento σ_{DVM}: desviación estándar de la dirección de los vectores de movimiento. Esta medida ofrece un indicador de la coherencia que existe entre el movimiento de los objetos de la escena. El sistema visual humano es más sensible al movimiento de objetos que tienen direcciones coherentes. Es decir, el ojo humano puede percibir mejor las imperfecciones en el movimiento de objetos si éstos se mueven en direcciones similares (coherentes).
- Módulo medio normalizado con respecto a la coherencia de movimiento μ_{MVM}/σ_{DVM}: cociente entre el módulo medio y la desviación típica de la dirección de los vectores de movimiento.

La información que proporcionan los vectores de movimiento es un buen indicador de la complejidad de codificación de cada secuencia de vídeo. La utilización de estas medidas se puede ver en otros trabajos de la literatura, como [Yang et al., 2007], [Jin et al., 2007] y [Ou et al., 2011a].

Implementación de las medidas de complejidad y entrenamiento Las medidas SI, TI, ASI, ATI, ATI-Sobel y ATI-Sobel 2 han sido implementadas en Matlab, ya que las últimas versiones de dicho software ofrecen una API muy potente para la manipulación de vídeo, además de las conocidas facilidades que ofrece para llevar a cabo operaciones como convoluciones, filtrados, cálculo de estadísticos, etc.

Las medidas que involucran operaciones sobre información relacionada con vectores de movimiento se han escrito en C, de manera independiente al resto de medidas, utilizando las librerías ffmpeg [Bellard, 2014] y libav [Libav, 2014].

4.3.5. Evaluación del modelo

Una vez completado el entrenamiento de la red neuronal, en este apartado se presenta el rendimiento de la misma. Como se dijo anteriormente, se han utilizado dos algoritmos de entrenamiento (Levenberg-Marquardt y regularización bayesiana) obteniendo un resultado ligeramente superior con el algoritmo de regularización bayesiana. A continuación se presentan los resultados que se han obtenido con cada uno.

4.3.5.1. Levenberg-Marquardt

El rendimiento conseguido en el entrenamiento en términos de coeficiente de correlación de Pearson (R), utilizando el algoritmo Levenberg-Marquardt, se muestra en la figura 4.15.

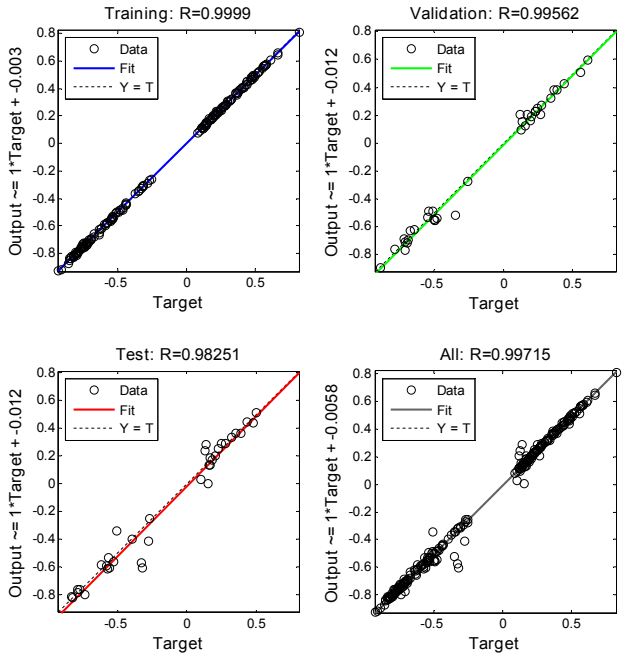


Figura 4.15: Rendimiento de la red neuronal entrenada con Levenberg-Marquardt

Como se puede ver, el ajuste con respecto a los datos de entrenamiento es excelente, y el error cometido en los datos de validación y test es también muy bueno, obteniendo un valor de $R > 0,98$.

Además de la correlación, se ha medido el error cuadrático medio MSE, obteniendo los siguientes resultados para cada uno de los sets de datos:

- MSE del set de entrenamiento: $6,43 \cdot 10^{-5}$
- MSE del set de validación: $0,002$
- MSE del set de test: $0,0074$

En la figura 4.16 se muestra la curva de aprendizaje en la que se destaca el punto de entrenamiento que se ha seleccionado. Dicho punto, de acuerdo a lo explicado en la sección anterior, es aquel en el que se alcanza un mínimo en el error asociado al conjunto de datos de validación, con el objetivo de evitar el overfitting. Este punto se alcanzó en la sexta iteración (epoch) del algoritmo de entrenamiento.

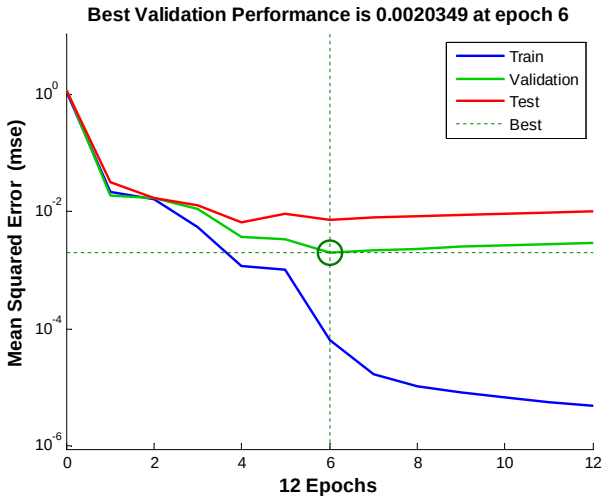


Figura 4.16: MSE de la red neuronal entrenada con Levenberg-Marquardt (curva de aprendizaje)

Es importante recordar que el error calculado en estas gráficas no mide diferencias entre valores de calidad sino las diferencias entre los valores esperados y estimados de los parámetros a y b del modelo.

A modo de ejemplo se incluyen algunas gráficas que comparan el valor de VQM_VFD esperado y el valor estimado por la red neuronal para algunos vídeos de prueba seleccionados al azar (no utilizados en el entrenamiento).

Como se puede ver en la figura y en la siguiente tabla, el error cometido para casi todos las tasas de bit es muy bajo, lo cual permite hacer estimaciones de VQM_VFD

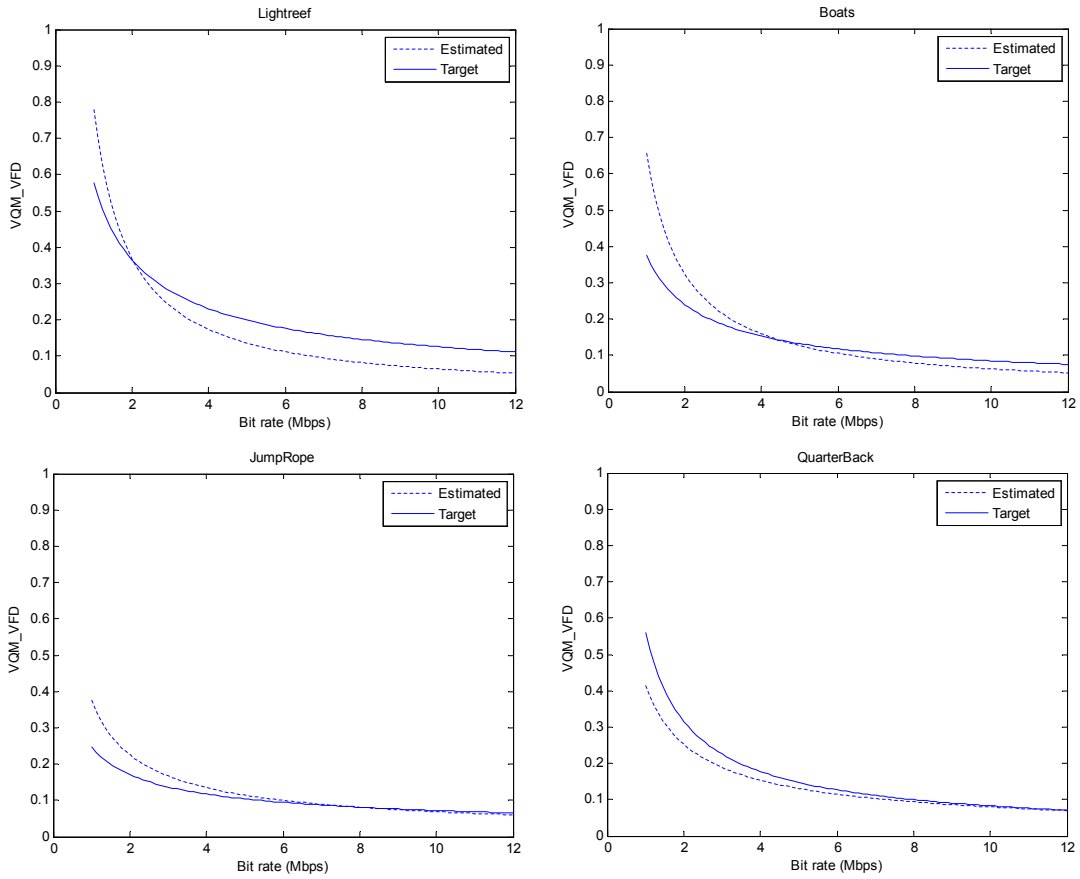


Figura 4.17: Estimación de VQM_VFD para secuencias de prueba no utilizadas en el entrenamiento (Levenberg-Marquardt)

muy precisas. En la menor tasa de bit considerada (1 Mbps) es donde se obtienen predicciones más burdas. Sin embargo, hay que poner en perspectiva estos resultados, teniendo en cuenta que para el tipo de servicios considerados en la tesis, tasas de bit tan bajas serán difícilmente aplicables, ya que resultan en valores de VQM_VFD mayores de 0,5, los cuales son inaceptables para la mayoría de servicios de vídeo sobre Internet.

Si se analiza el MSE, calculado sobre valores de VQM_VFD, de cada una de estas secuencias (con tasas de bit desde 1 Mbps a 12 Mbps) se obtienen los valores mostrados en la tabla 4.5, en la que el valor máximo del error es de $4,10 \cdot 10^{-3}$, es decir un 0,4 % del rango de VQM_VFD.

Si se cambia el rango de tasas de bit consideradas, considerando desde 2 Mbps a 12 Mbps, el MSE cometido mejora en todas las secuencias de vídeo de prueba, como se puede ver en la tabla 4.6. En este caso, el valor máximo del error es de $3,41 \cdot 10^{-3}$, es decir un 0,3 % del rango de VQM_VFD.

Tabla 4.5: MSE para secuencias de prueba no utilizadas en el entrenamiento (Levenberg-Marquardt). Tasa de bit de 1 a 12 Mbps

Secuencia de vídeo	**Error cuadrático medio**
Lightreef	$4,10 \cdot 10^{-3}$
Boats	$3,42 \cdot 10^{-3}$
JumpRope	$9,83 \cdot 10^{-4}$
QuarterBack	$1,35 \cdot 10^{-3}$

Tabla 4.6: MSE para secuencias de prueba no utilizadas en el entrenamiento (Levenberg-Marquardt). Tasa de bit de 2 a 12 Mbps

Secuencia de vídeo	**Error cuadrático medio**
Lightreef	$3,41 \cdot 10^{-3}$
Boats	$6,15 \cdot 10^{-4}$
JumpRope	$2,70 \cdot 10^{-4}$
QuarterBack	$4,24 \cdot 10^{-4}$

4.3.5.2. Regularización bayesiana

A continuación se presentan los resultados que se obtienen cuando el entrenamiento de la red neuronal se lleva a cabo utilizando el algoritmo de regularización bayesiana. En cuanto a la bondad del ajuste, los valores de correlación es muestran en la figura 4.18.

De manera similar al caso del entrenamiento Levenberg-Marquardt, el ajuste con respecto a los datos de entrenamiento es excelente, y el error cometido en los datos de test es también muy bueno, obteniendo un valor de $R > 0,99$.

Además de la correlación, se ha medido el error cuadrático medio MSE, obteniendo los siguientes resultados para cada uno de los sets de datos:

- MSE del set de entrenamiento: $1,69 \cdot 10^{-14}$
- MSE del set de test: $0,0041$

Se debe destacar que no se ha utilizado un conjunto de datos de validación, ya que la robustez de las redes neuronales entrenadas con regularización bayesiana así lo permite. La regularización bayesiana es un proceso matemático que convierte un problema de regresión no lineal en un problema estadístico bien definido en forma de regresión de arista (Regularización de Tíjonov o "ridge regression"). Este tipo de redes son difíciles de sobre-ajustar ya que la regularización bayesiana es capaz de calcular el número efectivo de parámetros de la red neuronal, desactivando aquellos que no sean relevantes.

Comparando estos resultados con el algoritmo anterior, el error cometido en el conjunto de test es algo menor con regularización bayesiana que con el algoritmo de

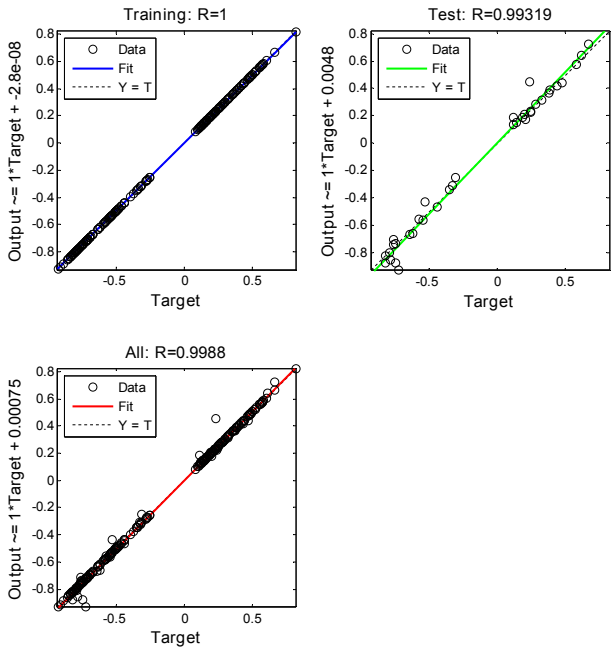


Figura 4.18: Rendimiento de la red neuronal entrenada con regularización bayesiana

Levenberg-Marquardt (0,0041 frente a 0,0074) . Sin embargo, la diferencia no es muy notable y se necesitarían más secuencias de vídeo de prueba para poder ofrecer una conclusión más firme.

De igual forma que se hizo con el algoritmo de entrenamiento Levenberg-Marquardt, a continuación se incluyen algunas gráficas que comparan el valor esperado y el valor estimado por la red neuronal para algunos vídeos de prueba (figura 4.19).

El MSE cometido en estas predicciones se presenta en la tabla 4.7. Como se puede ver, todos los valores obtenidos con regularización bayesiana mejoran el error cometido con el algoritmo Levenberg-Marquardt.

Tabla 4.7: Comparativa de algoritmos de entrenamiento en términos de MSE para secuencias de prueba no utilizadas en el entrenamiento. Tasa de bit de 1 a 12 Mbps

Secuencia de vídeo	**MSE R. bayesiana**	**MSE Levenberg-Marquardt**
Lightreef	$2,87 \cdot 10^{-3}$	$4,10 \cdot 10^{-3}$
Boats	$1,93 \cdot 10^{-3}$	$3,42 \cdot 10^{-3}$
JumpRope	$7,81 \cdot 10^{-4}$	$9,83 \cdot 10^{-4}$
QuarterBack	$7,71 \cdot 10^{-4}$	$1,35 \cdot 10^{-3}$

Al considerar tasas de bit desde 2 a 12 Mbps, el MSE cometido en el caso del entrenamiento con regularización bayesiana mejora en todas las secuencias de vídeo (en comparación con el caso anterior en el que se consideraban tasas de bit de 1 a

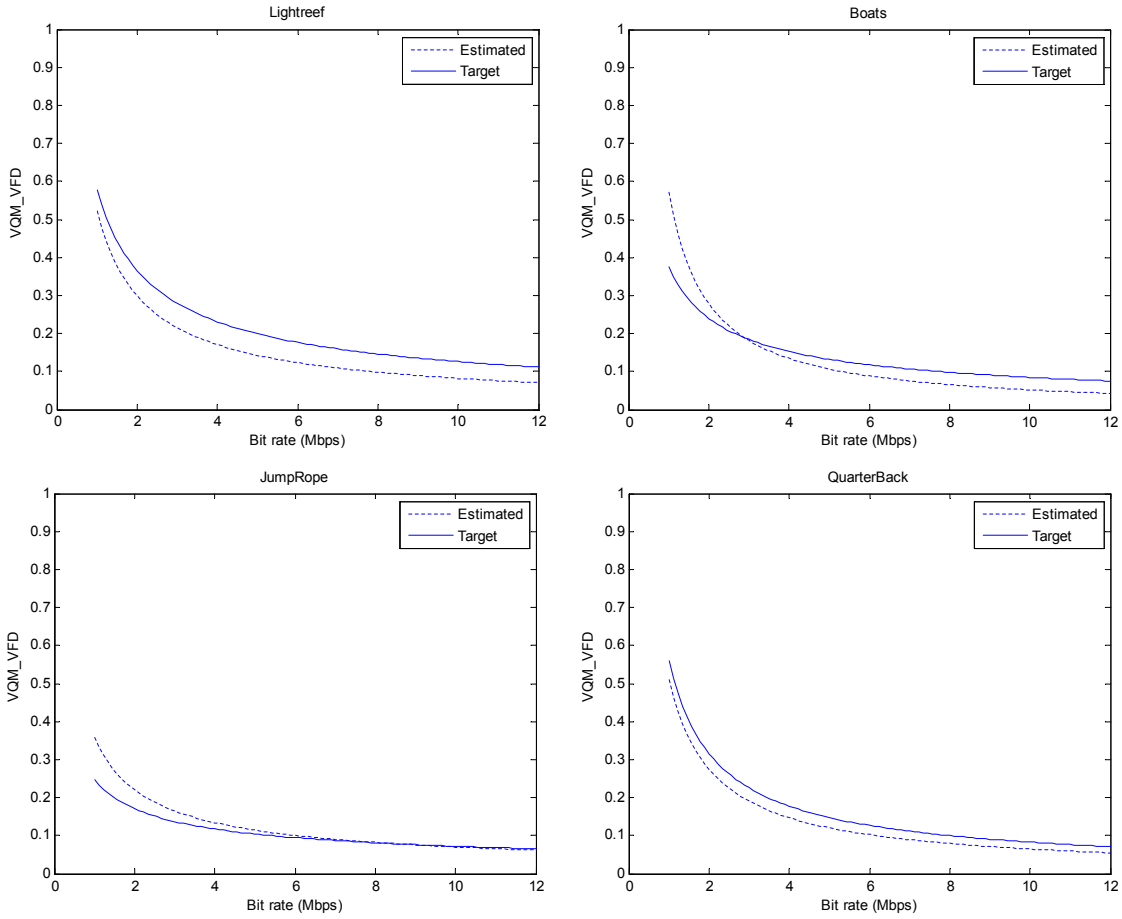


Figura 4.19: Estimación de VQM_VFD para secuencias de prueba no utilizadas en el entrenamiento (regularización bayesiana)

12 Mbps). Sin embargo, la regularización bayesiana no supera en todos los casos al algoritmo Levenberg-Marquardt al considerar un rango de tasas de bit de 2 a 12 Mbps (tabla 4.8).

4.4. Resumen y conclusiones

En este capítulo se ha desarrollado un modelo sin referencia para la estimación de la calidad percibida de vídeo, que cuantifica el efecto de las degradaciones de calidad asociadas al proceso de codificación. Este modelo es capaz de obtener una estimación de VQM_VFD (una métrica de calidad percibida de referencia completa) sin utilizar información de la secuencia de vídeo original.

Durante el desarrollo de este modelo, se ha analizado la relación entre la calidad percibida y la tasa de bit de codificación, concluyendo que ambas magnitudes se re-

Tabla 4.8: Comparativa de algoritmos de entrenamiento en términos de MSE para secuencias de prueba no utilizadas en el entrenamiento. Tasa de bit de 2 a 12 Mbps

Secuencia de vídeo	**MSE R. bayesiana**	**MSE Levenberg-Marquardt**
Lightreef	$2,75 \cdot 10^{-3}$	$3,41 \cdot 10^{-3}$
Boats	$8 \cdot 10^{-4}$	$6,15 \cdot 10^{-4}$
JumpRope	$2,28 \cdot 10^{-4}$	$2,70 \cdot 10^{-4}$
QuarterBack	$6,33 \cdot 10^{-4}$	$4,24 \cdot 10^{-4}$

lacionan mediante una función potencial. Además, se ha puesto de manifiesto que los parámetros de ajuste de esta función se pueden estimar mediante la utilización de un conjunto de medidas de complejidad espacial y temporal de la secuencia de vídeo. Más concretamente, estas medidas de complejidad de la secuencia de vídeo se utilizan como variables de entrada de una red neuronal, la cual, tras el entrenamiento pertinente, genera una estimación de los parámetros de ajuste de la curva VQM_VFD para la secuencia de vídeo que se está evaluando. Con el objetivo de conseguir un nivel adecuado de generalización en los resultados del modelo, se ha entrenado la red neuronal utilizando una amplia base de datos de secuencias de vídeo de prueba (VQEG HDTV).

Debido al bajo error obtenido en las estimaciones de VQM_VFD que genera el modelo propuesto y a su caracter sin referencia, este modelo puede ser utilizado para obtener en tiempo real una estimación de la calidad de vídeo en servicios de streaming OTT.

Capítulo 5

Modelo de degradación de calidad debida a la transmisión

5.1. Introducción

En el capítulo 3 se introdujo el factor I_{tra} cuya función es modelar la reducción en la calildad percibida que puede producirse como consecuencia de transmitir un flujo audiovisual a través de un canal TCP/IP de acuerdo al estándar MPEG-DASH, el cual puede introducir:

- Tiempos de espera e interrupciones en el contenido (buffering inicial y rebuffering).
- Variaciones en la calidad debidas al algoritmo de adaptación.

En este capítulo se desarrolla dicho factor, siguiendo una estructura similar al capítulo anterior. En primer lugar se lleva a cabo un estudio del estado del arte, revisando trabajos con objetivos alineados con el objetivo de este capítulo, los cuales sirven como base para la propuesta del modelo de degradación que se introduce a continuación.

5.2. Revisión del estado del arte

El análisis del estado del arte que se realiza en esta sección se va a dividir en dos subsecciones. La primera de ellas se centra en el estudio del efecto de los tiempos de espera e interrupciones en el contenido (buffering inicial y eventos de rebuffering). La segunda tiene como objetivo analizar los trabajos relacionados con la influencia de los cambios de calidad a lo largo del tiempo en la calidad percibida por los usuarios.

5.2.1. Buffering inicial y eventos de rebuffering

En [Tan et al., 2006] se estudia la calidad subjetiva del streaming de vídeo en entornos móviles donde las condiciones del medio dan lugar a eventos de buffering (buffering inicial y rebuffering). Los autores afirman que las degradaciones que mayor efecto suponen son la duración del evento de rebuffering y la frecuencia de rebuffering.

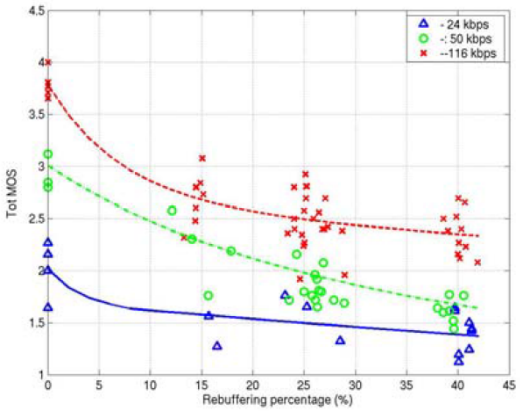


Figura 5.1: Calidad en función del tiempo de rebuffering. [Tan et al., 2006]

Además, afirman que si el rebuffering es inevitable, un solo evento de rebuffering es preferible a múltiples rebufferings más cortos. Es decir, dada una duración total de rebuffering, la degradación de la calidad es menor si solo hay un evento de rebuffering (ver figura 5.2).

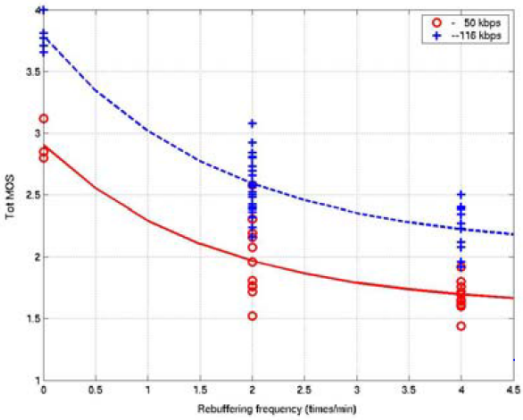


Figura 5.2: Calidad en función del número de eventos de rebuffering. [Tan et al., 2006]

Establecen también, de acuerdo a otros estudios analizados que los usuarios móviles toleran relativamente bien el buffering inicial, degradándose más la calidad cuando los eventos de rebuffering se producen en la parte final del contenido.

Aunque se llevan a cabo diferentes test subjetivos orientados a evaluar el efecto del buffering inicial y los eventos de rebuffering, no se plantea un modelo matemático para estimar dicha degradación a partir de parámetros objetivos y medibles.

En [Gustafsson et al., 2008] se puede encontrar una extensión del artículo anterior en el que los autores incluyen el efecto de las pérdidas de paquete. En esta versión del artículo los autores proponen la utilización de un modelo multiplicativo de estimación de QoE en función de la calidad de vídeo, las pérdidas y los eventos de rebuffering. Sin embargo, aunque presentan resultados de este modelo, no especifican formalmente el modelo desarrollado.

En [Mok et al., 2011] se propone un modelo de calidad percibida en servicios de streaming de vídeo sobre HTTP en función de parámetros de calidad de servicio de red. Más concretamente, los autores definen tres niveles de QoS:

- QoS de usuario o QoE.
- QoS de aplicación, donde se definen varias "métricas de rendimiento de aplicación":
  - Duración del buffering inicial.
  - Duración media de rebuffering.
  - Frecuencia de rebuffering.
- QoS de red, que incluye parámetros como Round-Trip delay Time (RTT), ancho de banda, tasa de pérdida de paquetes, etc.

A partir de la distinción de estos tres niveles de calidad, los autores establecen un modelo lineal que relaciona el nivel de QoE con el nivel de QoS de red, como se puede ver en la ecuación 5.1, donde L_{ti} es el nivel de de buffering inicial, L_{fr} es el nivel de frecuencia de rebuffering y L_{tr} es el nivel de duración media de rebuffering.

$$MOS = 4{,}23 - 0{,}0672 \cdot L_{ti} - 0{,}742 \cdot L_{fr} - 0{,}106 \cdot L_{tr} \tag{5.1}$$

Las variables L_{xy} toman valores de 1 a 3 en función del rango en el que se encuentre cada una de las variables, como se puede ver en la siguiente tabla:

Tabla 5.1: Niveles de degradación de QoE del modelo [Mok et al., 2011]

Nivel	**T. buffering inicial**	**F. rebuffering**	**T. medio de rebuffering**
1	0-1 segundos	0-0.02	0-5 segundos
2	1-5 segundos	0.02-0.15	5-10 segundos
3	>5 segundos	>0.15	>10 segundos

Basándose en este trabajo, en [Mok et al., 2012] los autores proponen un sistema consciente de la QoE para mejorar la calidad percibida por los usuarios de vídeo. Utilizan medidas del ancho de banda disponible para facilitar la selección de las distintas representaciones del vídeo. Han llevado a cabo pruebas subjetivas de QoE de las que se desprende que los usuarios prefieren cambios graduales en la calidad de las representaciones frente a cambios bruscos. Proponen también un algoritmo de adaptación para DASH basado en QoE.

El sistema que propone el artículo se basa en dos componentes: QDASH-abw and QDASH-qoe. QDASH-abw es una metodología de sondeo para detectar el nivel más alto de calidad que las condiciones de la red pueden soportar. Este módulo se implementa en forma de proxy capaz de medir el ancho de banda disponible por RTT. Con la estimación del ancho de banda que obtiene QDASH-abw, QDASH-qoe se encarga de ayudar a los clientes a seleccionar el nivel de calidad más adecuado, evitando saltos bruscos de niveles de calidad.

Es interesante destacar el hecho de que, según los autores, los usuarios muestran poca apreciación a las mejoras de calidad, mientras que critican fuertemente las degradaciones de calidad. Por tanto, es razonable pensar que los usuarios prefieren un nivel de calidad inicial más bajo frente a sufrir una gran degradación en la calidad cuando el throughput de la red decae.

En [Krishnan and Sitaraman, 2012] se establece una relación causal entre la calidad del vídeo y el comportamiento del espectador.

Los autores muestran que los espectadores empiezan a abandonar la visualización de un vídeo si este tarda más de 2 segundos en comenzar su reproducción, aumentando un 5.8 % la tasa de abandono por cada segundo extra de retardo inicial. Además, los usuarios son menos tolerantes al tiempo de buffering inicial en vídeos cortos que en vídeos largos como películas o series (ver figura 5.3). La probabilidad de que un espectador de vídeo corto abandone antes que un espectador de vídeo largo es un 11.5 % mayor de que ocurra al revés.

Por otro lado, los usuarios que acceden a Internet a través conexiones más rápidas tienen menos paciencia en cuanto al tiempo de buffering inicial y abandonan antes la reproducción del vídeo que usuarios con conexiones más lentas (ver figura 5.4). Más concretamente, los usuarios de vídeo en móvil son los más pacientes y los que menos abandonan, mientras que los que tienen conexiones de fibra óptica son los primeros en abandonar. La probabilidad de que un usuario de fibra abandone antes que un usuario móvil es un 38.25 % mayor de que ocurra al revés.

En cuanto al rebuffering, establecen que los usuarios que sufren un tiempo de rebuffering del 1 % de la duración del vídeo reproducen un 5 % menos de vídeo que un usuario que no haya experimentado rebuffering (ver figura 5.5). Por otra parte, los

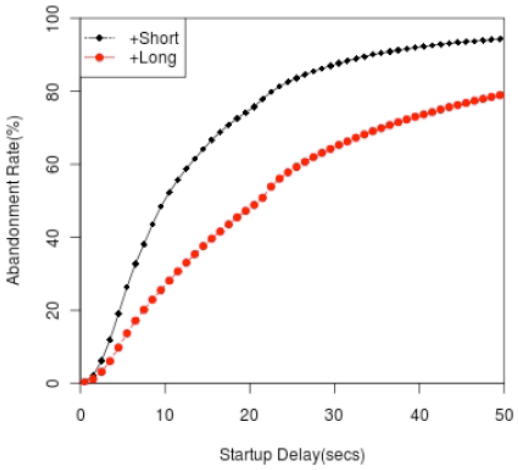


Figura 5.3: Tasa de abandono en función del tiempo de buffering inicial para diferentes duraciones de vídeo. [Krishnan and Sitaraman, 2012]

usuarios que sufren un fallo en el servicio son un 2.32 % menos propensos a reutilizar el servicio que aquellos usuarios que no hayan sufrido fallos.

En [Eckert et al., 2013] se propone un método de estimación de QoE para vídeo de descarga progresiva a partir de parámetros de red. Dicho método, denominado QMON (Quality Monitoring) consiste en realizar una estimación del nivel de buffer de cliente basándose en observaciones de los flujos TCP y evaluando la QoE en función del número y la duración de los eventos de rebuffering. El modelo tiene en cuenta también el tiempo transcurrido entre el último evento de rebuffering y el instante actual, con el objetivo de modelar un cierto efecto memoria. Con estos parámetros, los autores proponen el siguiente modelo, donde NI, denominado "impacto negativo" es un factor que combina los efectos de los tres parámetros considerados en el modelo:

$$MOS = 4{,}5 - NI;\ 0 \leq NI \leq 4{,}5 \tag{5.2}$$

$$NI = D_1 + D_2 - D_3 \tag{5.3}$$

El efecto del número de eventos de rebuffering viene dado por la ecuación 5.4, donde x representa el número de eventos de rebuffering y a es un parámetro de ajuste del modelo.

$$D_1 = e^{\frac{x}{a}} - 1 \tag{5.4}$$

El efecto del tiempo de rebuffering se modela de acuerdo a la ecuación 5.5, donde

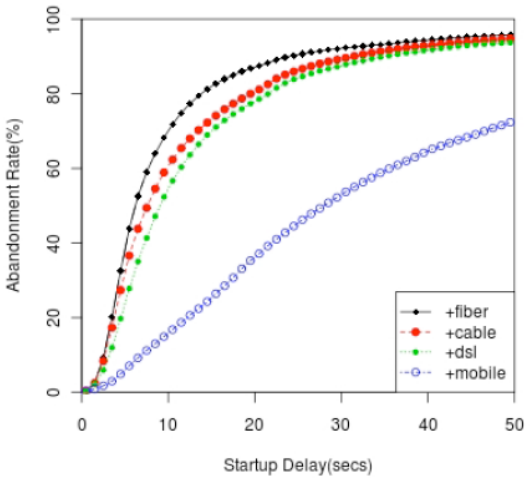


Figura 5.4: Tasa de abandono en función del tiempo de buffering inicial para distintas tecnologías de red de acceso. [Krishnan and Sitaraman, 2012]

t_{stall} es el tiempo de rebuffering y d es un parámetro de ajuste del modelo.

$$D_2 = e^{\frac{t_{stall}}{d}} - 1 \tag{5.5}$$

Los autores afirman que el valor de MOS se va incrementando progresivamente tras sufrir un evento de rebuffering, por lo que incluyen un factor positivo para modelar este efecto memoria. Este factor se define en la ecuación 5.6, donde t_{play} es el tiempo transcurrido desde el último evento de rebuffering y f es un parámetro de ajuste.

$$D_3 = \frac{f}{x} \cdot \sqrt[e]{t_{play}}, \forall x \in N \setminus \{0\} \tag{5.6}$$

Este efecto memoria obliga a reescribir el término D_1 para reducir el efecto del rebuffering al aumentar el tiempo desde el último evento de rebuffering (ecuaciones 5.7 y 5.8).

$$D_1 = e^{\frac{x}{a} - D_{11}} - 1 \tag{5.7}$$

$$D_{11} = b \cdot ln(t_{play} + 1) \tag{5.8}$$

En [Oyman and Singh, 2012] se analizan las métricas y mecanismos de reporte de QoE especificados en el estandar DASH del 3GPP. Este framework de monitorización de QoE permite al servidor solicitar a los clientes que lo soporten el envío de una serie de métricas de calidad:

- Logs de peticiones y respuestas HTTP

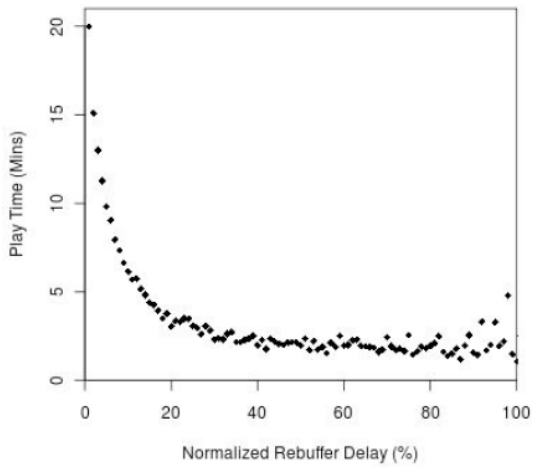


Figura 5.5: Tiempo de reproducción en función del tiempo de rebuffering. [Krishnan and Sitaraman, 2012]

- Lista de cambios de representación (cambios de calidad)
- Throughput medio
- Tiempo de buffering inicial
- Nivel del buffer
- Logs sobre el control de reproducción (pausas, accesos aleatorios, etc.)
- Información del MPD

El estudio realizado en [De Pessemier et al., 2013] tiene como objetivo investigar la influencia de los eventos de rebuffering en la calidad percibida en servicios de vídeo en dispositivos móviles. Dicho estudio se realizó mediante el diseño de seis escenarios de evaluación subjetiva de calidad que combinan tres tipos de conexiones móviles y dos tipos de calidades de vídeo.

Los resultados obtenidos muestran que aunque las interrupciones debidas a los eventos de rebuffering son molestas para los usuarios, éstos suelen aceptar un número limitado de ellas (en entornos móviles). En general, los usuarios prefirieron la reproducción continua (sin eventos de rebuffering) a un vídeo de mayor resolución, bitrate y tasa de frames. El tiempo de buffering inicial también se consideró menos importante que los eventos de rebuffering en cuanto a calidad percibida se refiere.

Con estos resultados los autores desarrollaron un modelo orientado a predecir la aceptabilidad de una sesión de vídeo en un dispositivo móvil, en función del tiempo de rebuffering. Según este modelo, las sesiones con un tiempo de rebuffering menor de 20 segundos tienen una probabilidad mayor que 0,75 de ser aceptadas por los usuarios,

mientras que si se experimenta un tiempo de rebuffering de más de 60 segundos, la probabilidad de que dicha sesión no sea aceptable para los usuarios es de más del 75 %.

En [Hossfeld et al., 2012] se evalúa el compromiso entre el tiempo de buffering inicial y el tiempo de rebuffering, llevando a cabo una serie de experimentos de evaluación de calidad subjetiva utilizando vídeos de Youtube. Según sus resultados, los eventos de rebuffering deben evitarse en cualquier caso, incluso a costa de incrementar el tiempo de buffering inicial.

En [Singh et al., 2012] se propone un método de monitorización de calidad para servicios de vídeo que utilicen streaming adaptativo sobre HTTP y codificación H.264. Las variables utilizadas para realizar dicha monitorización de calidad son el parámetro de cuantificación (QP) de H.264 y las interrupciones en la reproducción debidas a eventos de rebuffering. Los autores entrenan un modelo basado en redes neuronales aleatorias (RNN) y concluyen que los usuarios son más sensibles a los eventos de rebuffering que al incremento de QP para valores bajos del mismo. Cuando QP aumenta, la caída de la QoE es baja. Solo tras alcanzar un cierto valor de QP la QoE empieza a decaer más rápidamente.

En [Akhshabi et al., 2011] se describen y comparan los algoritmos de adaptación que aplican los servicios más utilizados actualmente (Microsoft SS, HLS, Netflix, etc.).

En general, el comportamiento habitual de los reproductores es seleccionar un nivel de calidad que tenga una tasa de bit menor que el througput medido, de manera que la velocidad de descarga es mayor que la velocidad de reproducción. Esto evita que haya eventos de rebuffering. Sin embargo, existen diferencias entre las implementaciones de distintos reproductores de vídeo. Por ejemplo, el reproductor de Netflix es más agresivo a la hora de intentar reproducir mayores niveles de calidad, mientras que el reproductor de Microsoft Smooth Streaming es más conservador.

5.2.1.1. Conclusiones extraídas del estado del arte

El análisis del estado de arte pone de manifiesto el creciente interés en la evaluación y estimación de la calidad percibida en sistemas de streaming de vídeo adaptativo sobre HTTP.

Como se ha podido ver, las principales variables o factores que influyen en la calidad (además de la propia calidad de vídeo, desde el punto de vista de la codificación) son:

- Tiempo de buffering inicial
- Número o frecuencia de eventos de rebuffering
- Tiempo de rebuffering

De estos factores, la mayoría de artículos consultados coinciden en destacar el efecto de los eventos de rebuffering como la componente que más condiciona la calidad percibida.

Además, se ponen de manifiesto algunas características del comportamiento humano muy interesantes, como por ejemplo que dado un tiempo de rebuffering, los usuarios aceptan mejor que dicho tiempo se concentre en un solo evento de rebuffering en vez de en varios eventos de rebuffering repartidos a lo largo del vídeo. También se pueden encontrar experimentos donde se compara el efecto que tiene que el evento de rebuffering se lleve a cabo al principio o al final del vídeo, demostrando que si el evento de rebuffering se produce al final del vídeo, la degradación de la calidad es mayor. Esto es así debido al efecto memoria del usuario, el cual hace que tras un cierto tiempo de reproducción fluida, el usuario asuma que no van a existir problemas de reproducción futuros.

Aunque en el estado del arte se pueden encontrar conclusiones muy interesantes, en los trabajos analizados no se formaliza matemáticamente de manera completa ningún modelo de calidad percibida, o bien se hace utilizando una métrica distinta a la contemplada en esta tesis.

Por otro lado, la metodología seguida en el desarrollo de algunos modelos de la literatura no garantiza la fiabilidad de los resultados conseguidos.

En el caso de [Hossfeld et al., 2011] y [Hossfeld et al., 2012] el diseño de los experimentos subjetivos llevados a cabo no garantiza que los resultados obtenidos no estén influidos por la calidad de vídeo y el propio contenido de las secuencias de prueba utilizadas.

Así pues, los resultados extraídos del estado del arte no serán utilizados directamente sino que se utilizarán algunas ideas para plantear un nuevo modelo capaz de formalizarlas matemáticamente.

5.2.2. Adaptación del nivel de calidad

En [Cranley et al., 2006] se aborda el problema de cómo adaptar la calidad de vídeo en términos de parámetros de codificación (resolución y tasa de frames) y calidad percibida en servicios de streaming de vídeo sobre redes IP best-effort.

Según este artículo, la mayoría de algoritmos de adaptación indican cómo se debe ajustar el bitrate del vídeo como respuesta a los cambios en las condiciones de la red. Sin embargo, esta adaptación no se suele plantear en términos de calidad percibida, ya que para conseguir un vídeo con una determinada tasa de bit, existen diferentes parámetros que pueden ser modificados (resolución, tasa de frames, etc.). Así pues, la hipótesis de los autores es que existe una trayectoria de adaptación óptima (OAT, del inglés Optimal Adaptation Trajectory), que maximiza la calidad percibida por el

usuario. Es decir, considerando el conjunto de posibles formas de obtener un bit rate objetivo, existe un conjunto de parámetros de codificación que maximiza la calidad percibida.

En la figura 5.6 se muestra la OAT, obtenida mediante tests subjetivos para 4 secuencias de vídeo con características espaciales y temporales diferentes. Dichas secuencias fueron extraídas de la base de datos del VQEG en formato YUV y codificadas en MPEG-4 con una tasa de frames máxima de 25 fps y una resolución máxima de 176x144 (QCIF). Los tests consistieron en definir regiones en el espacio resolución-frame rate que tuvieran la misma tasa de bit (zonas EABR, Equal Average Bit Rate) y pedir a los usuarios que evaluasen varias versiones de vídeos pertenecientes a la misma EABR. Como se puede ver en la figura, los test subjetivos sugieren que la OAT depende del tipo de contenido. En contenido con mucha acción (secuencias C1 y C2), la resolución es menos dominante, independientemente de las características espaciales de la secuencia. Esto implica que el usuario es más sensible a la fluidez del movimiento cuando hay mucha información temporal en la secuencia de vídeo. Análogamente, en las secuencias C3 y C4, las cuales tienen menos información temporal, la resolución se posiciona como el parámetro dominante.

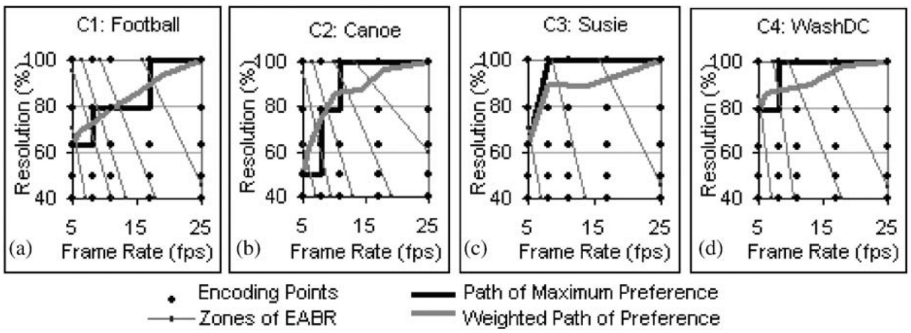


Figura 5.6: Trayectoria de adaptación óptima para distintos tipos de contenido. [Cranley et al., 2006]

Además, los autores afirman que utilizando una estrategia de adaptación basada en dos dimensiones (en el caso del artículo, frame rate y resolución), se consiguen mejores resultados que si la adaptación se realiza utilizando un único parámetro.

El artículo también estudia la posibilidad de obtener la OAT utilizando métricas objetivas, en concreto, utilizando VQM. Sin embargo, debido a los parámetros utilizados (frame rate y resolución), VQM no ofrece los mismos resultados en cuanto a OAT que los experimentos subjetivos. Este tipo de métricas objetivas están diseñadas sobre todo para evaluar degradaciones introducidas en la cadena de transmisión y no tanto para evaluar diferencias entre secuencias de vídeo con distinta resolución y frame rate. Estas métricas suelen basarse en la comparación pixel a pixel de las dos secuencias, por lo

que si las secuencias tienen una tasa de frames distinta, los resultados tienden a ser pobres.

En este artículo se presentan también los resultados de otro conjunto de experimentos subjetivos en los que se concluye que los participantes percibieron, en general, la adaptación del bit rate mediante la variación de la tasa de frames como el peor mecanismo de adaptación, seguida de la adaptación basada en resolución y de la adaptación OAT. Además, se observó que existe un retardo de varios segundos en la reacción de los participantes a los cambios de calidad. Por otro lado, la percepción de la adaptación de calidad es asimétrica, es decir, los usuarios son más críticos con la reducción de calidad y valoran menos de lo que cabría esperar el aumento de calidad.

Los mismos autores de este trabajo, en [Cranley et al., 2007] expanden el trabajo anterior y proponen un modelo objetivo para estimar la OAT (ecuación 5.9).

$$R = A \cdot (W \cdot ln(F) - (W - 1) \cdot ln(F_{max})) \tag{5.9}$$

Como se puede ver en la figura 5.7, en este modelo se ha incluido un factor W con el objetivo de dar más importancia a una de las dos componentes. Así pues, si $W > 1$, la componente dominante es la tasa de frames, mientras que si $W < 1$, la componente dominante es la resolución.

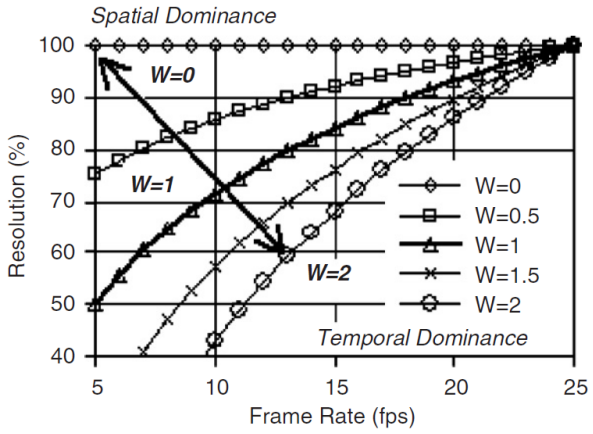


Figura 5.7: Modelo de estimación de trayectoria de adaptación óptima. [Cranley et al., 2007]

En [Gouache et al., 2011], para intentar evitar los cambios bruscos de calidad, especialmente molestos en contenidos HD, los autores de este artículo proponen utilizar simultáneamente varios servidores de vídeo, con el objetivo de contrarrestar la congestión de red y suavizar los cambios de calidad. Este enfoque se basa en combinar streaming adaptativo sobre HTTP con la técnica conocida como streaming distribuido [Nguyen and Zakhor, 2004]. La implementación que proponen los autores se basa en

utilizar la cabecera HTTP Range, para solicitar a cada servidor una porción de un fragmento de vídeo. La longitud de la porción solicitada, será proporcional a la estimación de ancho de banda de cada una de las rutas entre el cliente y los servidores.

En [Pinson and Wolf, 2003] se lleva a cabo una comparativa entre distintas metodologías subjetivas de evaluación de calidad de vídeo. Además, se analiza lo que los autores denominan "efecto memoria" en los resultados obtenidos mediante la metodología SSCQE (Single Stimulus Continous Quality Evaluation). En esta metodología, los evaluadores pueden puntuar de manera dinámica la calidad percibida mediante un selector asociado a una escala de calidad. El análisis del efecto memoria trata de responder a la pregunta de en qué medida la evaluación de calidad de los usuarios depende de las degradaciones que se produjeron a lo largo de la visualización del vídeo.

En primer lugar, los autores afirman que hay evidencias que afirman que los usuarios tienen memoria asimétrica, es decir, los usuarios penalizan rápidamente las degradaciones, pero no recompensan tan rápidamente las mejoras en la calidad.

En segundo lugar, afirman que los usuarios suele necesitar entre 9 y 15 segundos para formar su evaluación de calidad en los experimentos SSCQE.

En [Balachandran et al., 2013] se propone un modelo de estimación de QoE para vídeo sobre Internet con el objetivo de predecir no un valor de MOS sino una medida del involucramiento del usuario en el servicio (user engagement). Los autores afirman que este enfoque permite evaluar los sistemas de distribución de vídeo con unas métricas más afines a la tasa de retorno mediante publicidad y subscripciones de usuarios.

En el desarrollo de este modelo, los autores utilizan como métricas la tasa de bit media, el tiempo de buffering inicial, el porcentaje de rebuffering y la frecuencia de rebuffering. Aunque en este trabajo los autores no incluyen el efecto de la tasa de cambio de bit rate, en un trabajo anterior hacen referencia a dicho efecto. Más concretamente, en [Balachandran et al., 2012] los autores establecen que si la tasa de cambio de bit rate es menor que 0,5 cambios/minuto no hay efecto en el involucramiento del usuario, como se puede ver en la figura 5.8. Los datos presentados se han obtenido en medidas reales de usuarios que accedieron a portales de contenidos de vídeo, uno de ellos de series de TV y otro de eventos deportivos.

En [Zink et al., 2003] se lleva a cabo una evaluación subjetiva del efecto que tienen las variaciones de calidad en vídeos codificados en capas, llegando a conclusiones similares a las de otros trabajos. En primer lugar, la frecuencia de las variaciones debe ser la menor posible, y en segundo lugar, si no se puede evitar una variación, ésta debe ser lo menor posible. Sin embargo, en este artículo no se propone un modelo objetivo que permita evaluar el efecto que el cambio de calidad tiene en la calidad percibida por el usuario.

En [Mok et al., 2012] los autores proponen un mecanismo de adaptación de calidad

Figura 5.8: Involucramiento en función de la frecuencia de cambios de calidad. [Balachandran et al., 2012]

para DASH con el objetivo de mejorar la calidad percibida. Su propuesta se fundamenta en dos pilares: por un lado una arquitectura de proxys con los que obtener una estimación del ancho de banda más precisa que las generadas típicamente por los clientes DASH (QDASH-abw) y por otro lado, un algoritmo de selección de calidad donde se tiene en cuenta la calidad percibida a la hora de conmutar entre diferentes niveles de calidad (QDASH-qoe).

Básicamente, el algoritmo de selección de calidad que proponen los autores, se basa en evitar bajadas bruscas en el nivel de calidad. Es decir, apoyándose en el buffer del cliente, el algoritmo trata de disminuir la calidad de manera progresiva, calculando el número de fragmentos de calidad intermedia que puede solicitar hasta alcanzar el nivel de calidad objetivo (proporcionado por el módulo QDASH-abw).

Para el diseño de este algoritmo de adaptación, los autores de este artículo llevaron a cabo una serie de experimentos subjetivos en los que intentaron evaluar el efecto que tiene el cambio de nivel de calidad en la calidad percibida. Más concretamente, en los experimentos realizados se evaluó la calidad percibida por los usuarios al realizar una conmutación desde un nivel de calidad correspondiente a vídeo codificado a aproximadamente 4 Mbps a un nivel de calidad correspondiente a vídeo codificado a 400 Kbps, pasando por varios niveles intermedios (diferentes en cada experimento).

De entre los resultados que ofrece este artículo, destaca el hecho de que vídeos con una calidad de codificación media menor que otros, obtienen mayor MOS debido a la configuración de adaptación de calidad que se ha realizado. Sorprende especialmente el experimento en el que se compara una secuencia de vídeo formada por 14 segundos a

máxima calidad (nivel 4) y 9 segundos a mínima calidad (nivel 0) frente a otra secuencia formada por 10 segundos a nivel 3 de calidad y 13 segundos a nivel 1 de calidad, siendo el bitrate medio de estas secuencias de 2200 Kbps y 1000 Kbps respectivamente. Según los autores, la diferencia entre la MOS de la segunda secuencia y de la primera es de 1.24.

Este ejemplo es uno de los resultados más extremos, y puede ser explicado si se tiene en cuenta que, en primer lugar, la segunda secuencia no alcanza el nivel de calidad mínimo, y en segundo lugar, al no comenzar dicha secuencia con un valor de calidad muy alto, los cambios en la misma no son tan abruptos.

En general, la conclusión que se puede extraer de este artículo es que el efecto de la adaptación de calidad es proporcional a la diferencia de calidad de los niveles entre los que se conmuta. Para los intereses de esta tesis, aunque en este artículo se evalúa la diferencia en términos de MOS para distintas estrategias de adaptación, los resultados que se ofrecen no se muestran en su totalidad (solo se muestra una tabla en la que se indica la diferencia de MOS para algunas secuencias de vídeo seleccionadas), por lo que no se puede derivar un modelo lo suficientemente preciso a partir de los experimentos realizados en este artículo.

5.2.2.1. Conclusiones extraídas del estado del arte

Como se desprende del análisis de los trabajos relacionados, el estudio que el efecto de la adaptación de calidad tiene en la calidad percibida por los usuarios es un área que suscita interés entre la comunidad científica. Sin embargo, el enfoque típico que se ha aplicado para abordar este tema es algo distinto al que se aplica en esta tesis. En general, todos los trabajos identifican la importancia que la adaptación de calidad tiene en la calidad percibida, pero se centran en desarrollar mecanismos que permitan realizar la adaptación de la calidad de manera sensible a las percepciones del usuario. Por el contrario, como ya se ha comentado, esta tesis doctoral está centrada en evaluar de manera objetiva el efecto que las distintas componentes del servicio de streaming de vídeo OTT tienen sobre la QoE. Así pues, el objetivo concreto que se persigue en esta sección es poder cuantificar de manera objetiva el efecto que tiene el cambio de calidad en la MOS de los usuarios.

De los trabajos analizados, únicamente [Balachandran et al., 2013] trata de evaluar el efecto que producen los cambios de calidad en la percepción del usuario. Sin embargo, la métrica utilizada para evaluar dicho efecto (involucramiento de los usuarios) es distinta a la utilizada en esta tesis (MOS) y la conversión entre ambas no es directa.

En general, se pueden extraer una serie de ideas comunes que pueden servir como fundamento para el desarrollo de un nuevo modelo objetivo que cuantifique el efecto de la adaptación de calidad:

- La adaptación de calidad juega un papel importante en la opinión de un usuario acerca de una secuencia de vídeo, por lo que debe ser tenida en cuenta en un modelo global de estimación de QoE como el que se desarrolla en esta tesis.
- La degradación de la calidad percibida es proporcional al número de cambios en el nivel de calidad.
- La degradación de calidad percibida es proporcional a la diferencia de los niveles de calidad entre los que se conmuta.
- La percepción de los cambios de calidad es asimétrica: se penaliza más un cambio a un nivel inferior de calidad que lo que se premia un cambio a un nivel de calidad superior.
- La complejidad del contenido de la secuencia en la que se produce el cambio de calidad influye en la percepción del usuario.

5.3. Desarrollo del modelo

5.3.1. Introducción

En el capítulo 3 se presentó el modelo global de estimación de calidad para servicios de streaming de vídeo adaptativo. Más concretamente, en la ecuación 3.10 se introduce el factor I_{tra} que modela las degradaciones en la calidad percibida que introduce la red y los protocolos y mecanismos de transporte utilizados en este tipo de sistemas de streaming.

Los factores que modela la componente I_{tra} son los siguientes:

- Tiempo de buffering inicial
- Eventos de rebuffering
  - Tiempo total de rebuffering
  - Número de eventos de rebuffering
- Efecto de los mecanismos de adaptación de calidad

La expresión general de I_{tra} se presenta en la ecuación 5.10.

$$I_{tra} = I_{Tbuffering\,inicial} + I_{Trebuffering} + I_{Nrebuffering} + I_{\Delta Q} \tag{5.10}$$

Como se puede ver, el modelo I_{tra} tiene como objetivo capturar todas las degradaciones que la red puede introducir en un sistema de streaming de vídeo adaptativo

sobre HTTP, distinguiendo entre el retardo de buffering inicial y los siguientes eventos de rebuffering. Se contempla también el efecto que tiene que el tiempo total de rebuffering se reparta entre distintos eventos de rebuffering. Por último, se incluye una componente cuyo objetvo es contemplar el efecto de las adaptaciones de calidad que pueden producirse a lo largo de la reproducción del contenido.

5.3.2. Metodología: experimentos de evaluación subjetiva de calidad de vídeo

Para desarrollar cada uno de los componentes de I_{tra} se ha seguido una metodología basada en experimentos de evaluación subjetiva de calidad de vídeo, como se puede ver en la figura 5.9. En primer lugar, para cada componente del modelo se ha diseñado y se ha llevado a cabo un experimento en el que varios evaluadores puntuaron la calidad percibida en un conjunto de vídeos sometidos a diferentes degradaciones controladas. Los resultados extraídos de estos experimentos han permitido plantear modelos matemáticos que capturan la opinión media de los usuarios al valorar diferentes tipos de degradaciones de calidad.

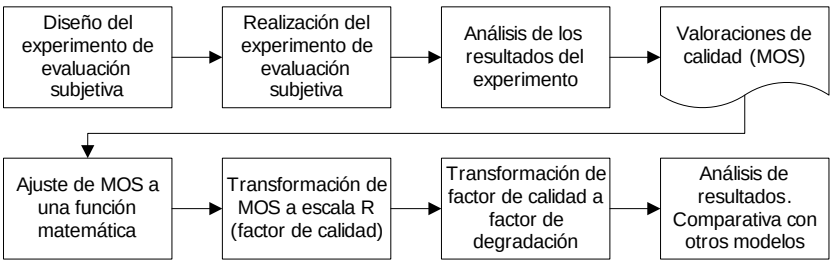


Figura 5.9: Metodología seguida en el desarrollo del modelo de degradación debida a la transmisión

La realización de tests de evaluación subjetiva de calidad es la mejor herramienta para conseguir datos fiables que permitan el desarrollo y la validación de modelos de estimación de calidad percibida. Tradicionalmente, estos tests se han llevado a cabo en laboratorios o salas especializadas en las que se emulaban las condiciones de visionado típicas de los sistemas considerados. Los evaluadores acudían a estas salas y llevaban a cabo el visionado de los contenidos y la valoración de calidad de los mismos. Existen diversas recomendaciones ITU que ofrecen guías sobre cómo llevar a cabo estos experimentos, orientando a los investigadores en aspectos que van desde el tamaño de la sala, distancias de visionado, hasta metodologías de evaluación y recogida de datos. Algunas de estas recomendaciones son: [ITU, 1997b], [ITU, 2012a], [ITU, 2008f].

Sin embargo, aunque esta filosofía de "entorno controlado" ofrece muy buenos resultados en cuanto a la calidad de las valoraciones recogidas, el coste asociado a la

realización de los experimentos es muy alto. Con esta motivación, en los últimos años ha surgido un nuevo paradigma de evaluación subjetiva de calidad basado en "crowdsourcing" [Chen et al., 2009], [Chen et al., 2010], [Xu et al., 2012], [Keimel et al., 2012], [Rainer et al., 2013], [Figuerola Salas et al., 2013].

La evaluación subjetiva de calidad basada en crowdsourcing consiste en externalizar la valoración de los contenidos a un conjunto de usuarios externos. En este nuevo contexto la valoración no se realiza en un laboratorio sino que se realiza online, es decir, cada evaluador accederá de manera remota desde su dispositivo a los contenidos a evaluar. En este tipo de tests los evaluadores se suelen conseguir, a cambio de una pequeña cantidad de dinero, mediante el uso de portales especializados en crowdsourcing, como Amazon Mechanical Turk [Amazon, 2014] o Microworkers [Microworkers, 2014].

A diferencia del enfoque tradicional, la realización de test subjetivos mediante crowdsourcing suele ser menos costosa, pero el control que se ejerce sobre el experimento es más reducido.

En esta tesis, se ha llevado a cabo un enfoque híbrido entre ambos paradigmas. En concreto la realización de la evaluación de calidad se lleva a cabo online mediante una plataforma de evaluación web de calidad de vídeo. Sin embargo, los usuarios que realizan el test son previamente seleccionados con el objetivo de reducir el porcentaje de datos falseados que pueden aparecer en test subjetivos realizados mediante crowdsourcing.

En el apéndice C se ofrecen más detalles sobre la plataforma web de evaluación de calidad de vídeo que se ha desarrollado y utilizado en esta tesis.

En los experimentos de evaluación subjetiva de calidad se han utilizado diferentes secuencias de vídeo que forman parte de las siguientes clases de tipos de contenido:

- Noticias
- Trailers de películas
- Vídeos musicales
- Vídeos deportivos

En cuanto a la duración de cada uno de los vídeos, en la literatura son comunes las pruebas con vídeos de entre 30 y 60 segundos, por lo que se ha seguido la misma pauta.

La metodología que se ha elegido para los experimentos es ACR-HR (Absolute Category Rating with Hidden Reference) [ITU, 2008f] con una escala de calidad de 5 puntos (1: Malo, 2: Pobre, 3: Razonable, 4: Bueno, 5: Excelente). La metodología ACR-HR consiste en evaluar un conjunto de secuencias de vídeo de manera independiente, incluyendo en dicho conjunto una versión de referencia (sin degradaciones) de cada una de las secuencias de vídeo de prueba (referencia oculta).

La principal ventaja de la utilización de la referencia oculta (frente a la metodología ACR) es que el impacto perceptual del vídeo de referencia puede ser eliminado

de las valoraciones subjetivas. Esto reduce la desviación asociada al contenido (ciertos contenidos gustarán más a los usuarios que otros), a la calidad de la señal de referencia (artefactos de codificación), y otros factores. Así pues, la utilización de vídeos de referencia oculta hace posible aislar el efecto concreto que se quiere estudiar en cada experimento.

Una vez introducida la metodología seguida, a continuación se detalla cada una de los componentes de I_{tra}.

5.3.3. Tiempo de buffering inicial

El experimento de evaluación subjetiva de calidad que se realizó para obtener datos sobre la degradación asociada al tiempo de buffering inicial incluyó un conjunto de vídeos de prueba con los siguientes tiempos de buffering inicial:

- $T_{buffering\,inicial} = 0\,s$ (secuencia de referencia)
- $T_{buffering\,inicial} = 2\,s$
- $T_{buffering\,inicial} = 10\,s$
- $T_{buffering\,inicial} = 25\,s$

Los resultados obtenidos en el experimento y el modelo propuesto se pueden ver en la figura 5.10. Se debe destacar que los puntos que se representan en la figura representan la degradación media para los distintos tipos de contenido considerados en los experimentos.

Como se puede ver, la degradación de la calidad crece de forma moderada con el tiempo de buffering inicial. Los datos subjetivos obtenidos en el experimento se han ajustado numéricamente a la siguiente curva:

$$I_{Tbuffering\,inicial} = a \cdot \sqrt{b \cdot T_{buffering\,inicial}} + c \tag{5.11}$$

Los parámetros de ajuste del modelo (a, b y c) toman los valores que se muestran en la tabla 5.2. Con estos valores, el ajuste consigue un coeficiente de correlación de 0,9925 y un RMSE de 3,268.

Tabla 5.2: Parámetros de ajuste del modelo de degradación asociada al tiempo de buffering inicial

a	b	c
3,611	7,957	-2,946

En la figura 5.11 se compara el modelo propuesto en esta tesis con otros modelos disponibles en la literatura que también estudian el efecto del tiempo de buffering inicial en la calidad percibida.

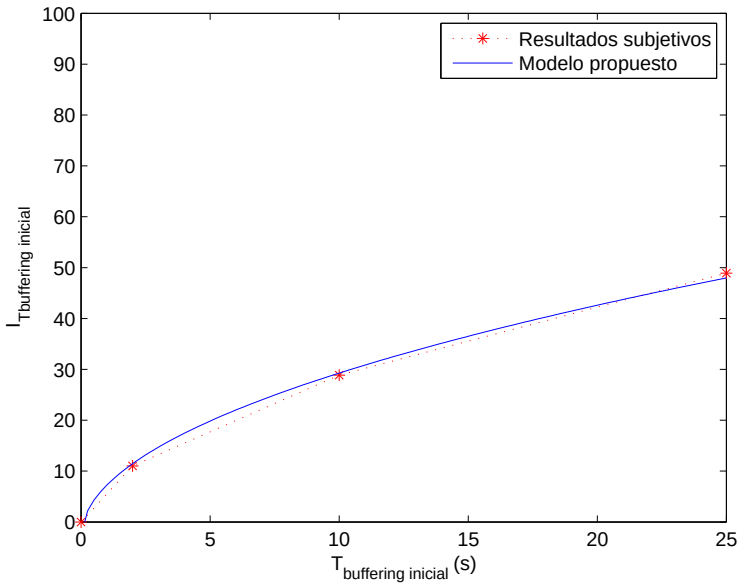


Figura 5.10: Efecto del tiempo de buffering inicial: valoraciones subjetivas y modelo propuesto

En primer lugar, se puede ver que la forma de curva obtenida es similar a la de otros modelos propuestos en la literatura. Sin embargo, aunque la forma es similar, los valores que proporcionan los distintos modelos se pueden agrupar en dos grupos. El primer grupo, formado por el modelo [Mok et al., 2011] y el modelo [Hossfeld et al., 2012], muestra una variación de la degradación con respecto al tiempo de buffering inicial muy moderada. El segundo grupo, en el que se encuentra el modelo propuesto en esta tesis y el modelo [Krishnan and Sitaraman, 2012], muestra una degradación de calidad mucho más agresiva con el aumento del tiempo de buffering inicial.

El motivo que con mayor probabilidad explica esta diversidad es la diferencia en las condiciones y en la metodología utilizadas para obtener los datos de valoración subjetiva en los que se basa cada modelo.

En cuanto a los modelos pertenecientes al denominado "primer grupo", [Mok et al., 2011] utiliza solo una secuencia de vídeo de prueba y el rango de valores de $T_{buffering\,inicial}$ es muy limitado (de 0 a 5 segundos). En [Hossfeld et al., 2012], como se comentó en el estado del arte, la no utilización de una señal de referencia puede hacer que el contenido influya en la valoración de calidad. Además, este modelo se centra en la evaluación de calidad de vídeos de Youtube, por lo que el carácter gratuito del servicio puede relajar las expectativas de los usuarios.

En el segundo grupo de modelos, [Krishnan and Sitaraman, 2012] utiliza una gran cantidad de datos subjetivos extraídos diversos servicios de vídeo tanto gratuitos como

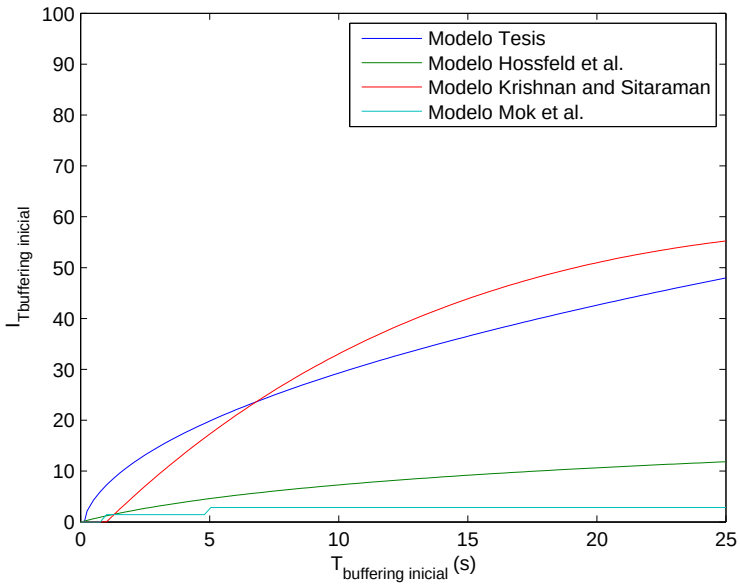


Figura 5.11: Efecto del tiempo de buffering inicial: comparativa con otros modelos

de pago que utilizan la CDN de Akamai, por lo que es de esperar que, en media, las valoraciones de calidad sean más exigentes. En cuanto al modelo propuesto en esta tesis, la utilización de referencia oculta reduce el efecto que puede tener el contenido en la valoración subjetiva de los usuarios.

5.3.4. Eventos de rebuffering

5.3.4.1. Tiempo de rebuffering

Para estudiar el efecto que tiene el tiempo de rebuffering en la calidad percibida, se diseñó un experimento de evaluación subjetiva de calidad en el que a cada evaluador se le mostró un conjunto de vídeos utilizando varios valores de $T_{rebuffering}$ para cada uno:

- $T_{rebuffering} = 0\,s$ (secuencia de referencia)
- $T_{rebuffering} = 2\,s$
- $T_{rebuffering} = 10\,s$
- $T_{rebuffering} = 25\,s$

Es importante destacar que el tiempo de rebuffering de cada vídeo se concentra en un único evento de rebuffering. El efecto que tiene el número de eventos de rebuffering se estudia en la siguiente sección.

Los resultados obtenidos en el experimento y el modelo propuesto se pueden ver en la figura 5.12.

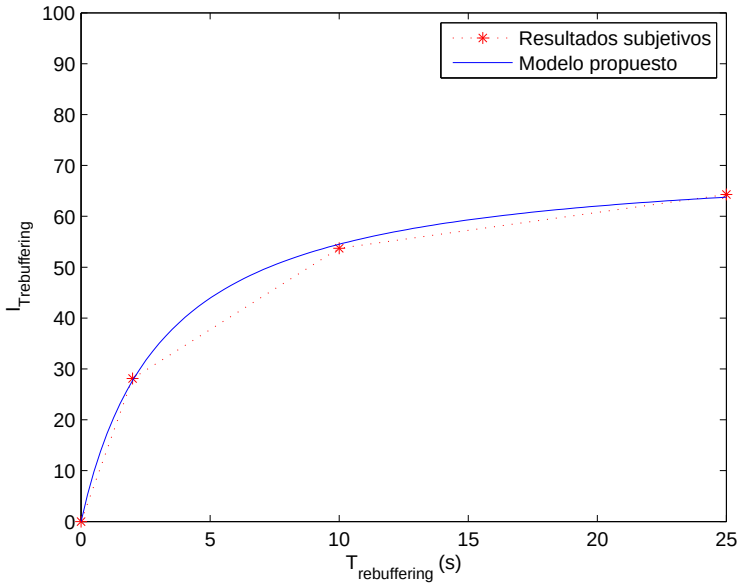


Figura 5.12: Efecto del tiempo de rebuffering: valoraciones subjetivas y modelo propuesto

Los resultados subjetivos extraídos del experimento se pueden ajustar a la hipérbola modificada que se presenta en la ecuación 5.12.

$$I_{Trebuffering} = \frac{a \cdot T_{rebuffering}}{1 + b \cdot T_{rebuffering}} \tag{5.12}$$

Cuando los parámetros de ajuste de la ecuación 5.12 toman los valores de la tabla 5.3, el coeficiente de correlación es de 0,9996 y el RMSE es de 0,73.

Tabla 5.3: Parámetros de ajuste del modelo de degradación asociada al tiempo de rebuffering

a	b
22,54	0,3135

Si se compara el efecto del tiempo de rebuffering con el efecto del tiempo de buffering inicial, se pone de manifiesto la mayor importancia del rebuffering en cuanto a la calidad percibida. Este resultado está en la línea con el trabajo de [Hossfeld et al., 2012], el cual analiza el compromiso entre tiempo de rebuffering y tiempo de buffering inicial, concluyendo que es preferible incrementar el tiempo de buffering inicial con el objetivo de disminuir el tiempo de rebuffering.

En la figura 5.13 se compara el modelo propuesto en esta tesis con otros modelos disponibles en la literatura.

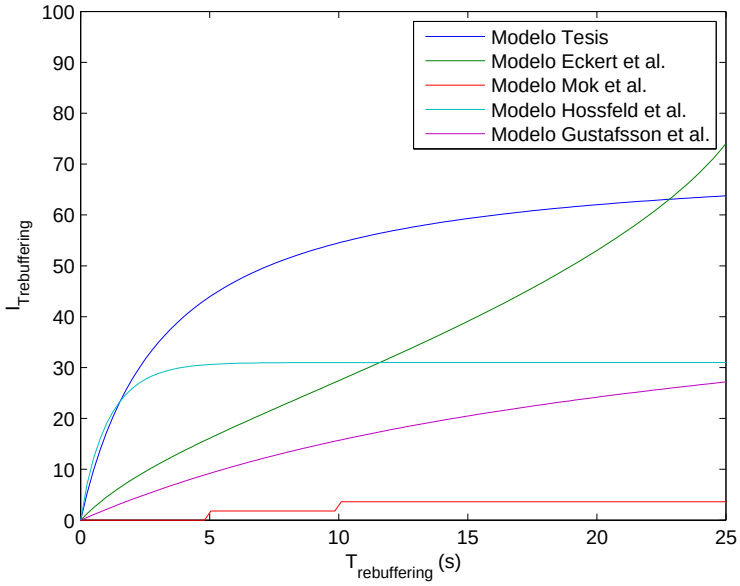


Figura 5.13: Efecto del tiempo de rebuffering: comparativa con otros modelos

Se puede apreciar cierta diversidad entre los modelos propuestos en la literatura, lo que de nuevo vuelve a poner de manifiesto las diferencias en las condiciones de evaluación de la calidad. El modelo propuesto en esta tesis utiliza una forma de curva similar a la de los modelos [Hossfeld et al., 2012] y [Gustafsson et al., 2008] pero la variación propuesta en la tesis es la más exigente en cuanto al tiempo de rebuffering. Se debe destacar que, para valores de $T_{rebuffering}$ bajos, el modelo de esta tesis y el modelo de [Gustafsson et al., 2008] son muy similares. El modelo de [Gustafsson et al., 2008], aunque algo antiguo, está desarrollado mediante tests subjetivos realizados con referencia oculta. Sin embargo, el rango de valores de $T_{rebuffering}$ considerado es muy bajo ($< 5s$), lo cual explica la saturación de dicho modelo a partir de $T_{rebuffering} = 5s$.

5.3.4.2. Número de eventos de rebuffering

Para analizar el efecto del número de eventos de rebuffering en la calidad percibida se llevó a cabo un experimento de evaluación subjetiva de calidad con los siguientes objetivos:

- Objetivo 1: dado un tiempo total de rebuffering, $T_{rebuffering}$, analizar cómo afecta que dicho tiempo se reparta entre varios eventos de rebuffering. Expresado

matemáticamente, se trata de encontrar la siguiente relación: $I_{Nrebuffering} = f(N_{rebuffering})$.

- Objetivo 2: analizar si el tiempo total de rebuffering, $T_{rebuffering}$, afecta a la degradación asociada al número de eventos de rebuffering. Expresado matemáticamente, se quiere evaluar si $I_{Nrebuffering} = f(N_{rebuffering}, T_{rebuffering})$.

Así pues, las degradaciones introducidas en las secuencias de vídeo de prueba utilizadas en el experimento de evaluación de calidad fueron las siguientes:

- Tiempo total de rebuffering:
  - $T_{rebuffering} = 6\,s$
  - $T_{rebuffering} = 10\,s$
- Número de eventos de rebuffering:
  - $N_{rebuffering} = 1$
  - $N_{rebuffering} = 2$
  - $N_{rebuffering} = 4$
  - $N_{rebuffering} = 6$

Una vez realizado el experimento y evaluados los resultados, en primer lugar se comprobó que la diferencia entre los resultados obtenidos en las secuencias de vídeo con $T_{rebuffering} = 6\,s$ y con $T_{rebuffering} = 10\,s$ es muy pequeña, como se puede ver en la figura 5.14.

Así pues, en cuanto al objetivo 2 planteado anteriormente, el modelo de $I_{Nrebuffering}$ será independiente de $T_{rebuffering}$.

En cuando al objetivo 1, utilizando el conjunto de valoraciones subjetiva conseguidas en el experimento se obtiene la gráfica de la figura 5.15, donde se presentan de manera conjunta los datos subjetivos y la curva propuesta. Como se puede ver, los resultados obtenidos confirman las ideas analizadas en el estado del arte que planteaban que dado un cierto tiempo de rebuffering es mejor (desde el punto de vista de la calidad percibida) si éste se concentra en un único evento de rebuffering.

Así pues, la ecuación que modela la degradación asociada al número de eventos de rebuffering es la siguiente:

$$I_{Nrebuffering} = a \cdot (1 - N_{rebuffering}^{b}) \tag{5.13}$$

Los parámetros de ajuste que se han utilizado para generar la figura 5.15 se recogen en la tabla 5.4. Con estos valores se obtiene un coeficiente de correlación de 0,994 y un RMSE de 1,329.

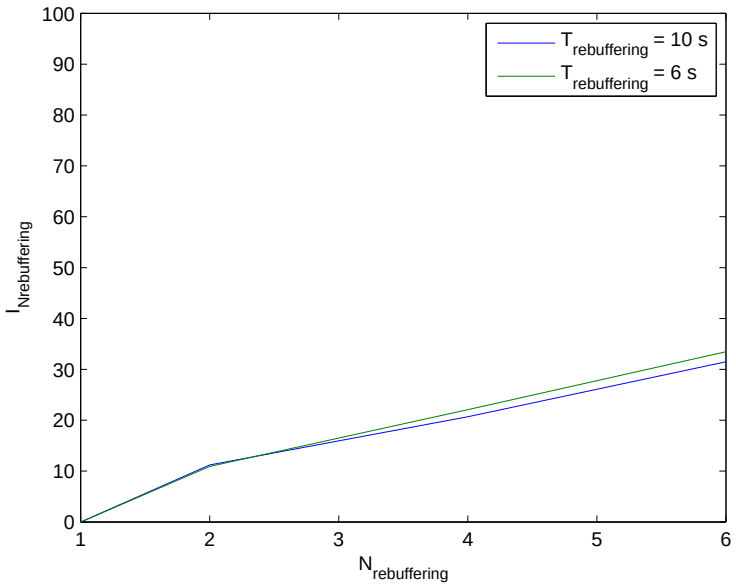


Figura 5.14: Efecto del número de eventos de rebuffering con respecto al tiempo total de rebuffering

Tabla 5.4: Parámetros de ajuste del modelo de degradación asociada al número de eventos de rebuffering

a	b
-30,43	0,4032

La comparación del modelo propuesto en esta tesis con otros modelos propuestos en la literatura se puede ver en la figura 5.16.

Como se puede observar, todos los modelos comparados siguen la misma tendencia y ofrecen valores relativamente similares. Los valores ofrecidos por el modelo de esta tesis están especialmente próximos a los del modelo de [Eckert et al., 2013]. Del resto de modelos, el menos exigente es el de [Tan et al., 2006]. El modelo de [Mok et al., 2011] es el que alcanza mayores niveles de degradación en ciertos puntos de la curva (valores de $N_{rebuffering} = 2$ y $N_{rebuffering} = 4$, lo cual es lógico teniendo en cuenta los pesos que dicho modelo da a cada componente de degradación (ver ecuación 5.1).

5.3.5. Adaptación de calidad de vídeo

Como ya se ha comentado, los algoritmos de adaptación de los clientes DASH pueden decidir conmutar entre diferentes niveles de calidad a lo largo de la reproducción de un contenido, como consecuencia de las condiciones cambiantes de la red, del dispositivo de visualización o de otros factores. Este comportamiento es compatible con el modelo

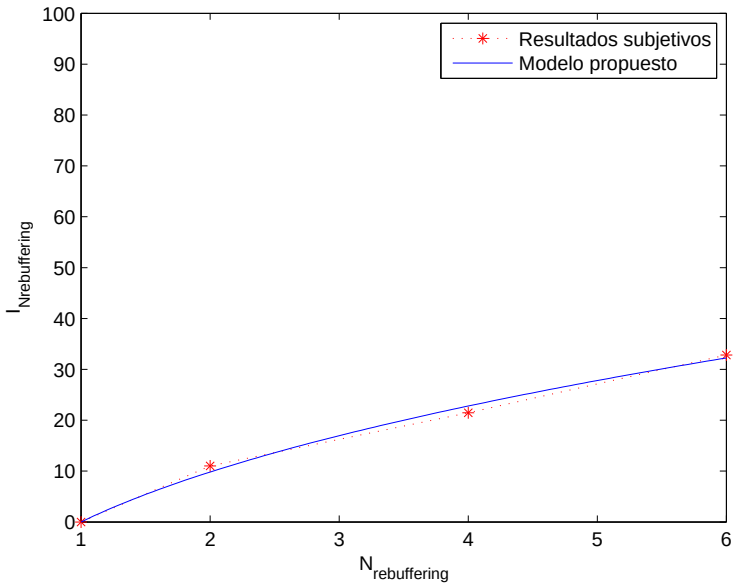


Figura 5.15: Efecto del número de eventos de rebuffering: valoraciones subjetivas y modelo propuesto

de estimación de calidad global propuesto en esta tesis, ya que el resultado generado por el modelo de calidad de vídeo reflejará los cambios de calidad que se produzcan. Sin embargo, como se pone de manifiesto en la revisión de la literatura que se llevó a cabo en la sección 5.2.2, los cambios en la calidad de vídeo pueden introducir una degradación adicional en la calidad percibida.

Para obtener un mayor conocimiento sobre esta degradación, se diseñó un experimento de evaluación subjetiva de calidad en el que se simularon conmutaciones entre distintos niveles de calidad de vídeo. En primer lugar, se seleccionaron un conjunto de vídeos, con diferentes complejidades espaciales y temporales. En segundo lugar, se definieron una serie de "trayectorias de adaptación" para cubrir diversos aspectos como:

- "Distancia" entre niveles de calidad: el concepto de distancia entre niveles mide la diferencia entre la calidad de dos representaciones del contenido sucesivas.
- "Sentido" (ascendente o descendente) del nivel de calidad: el sentido ascendente se refiere a una mejora en la calidad del contenido, mientras que el sentido descendente hace referencia a una bajada en el nivel de calidad.
- Número de cambios de calidad.

En tercer lugar, con el objetivo de hacer abordable el estudio de estas tres variables en un único experimento, se realizaron una serie de consideraciones:

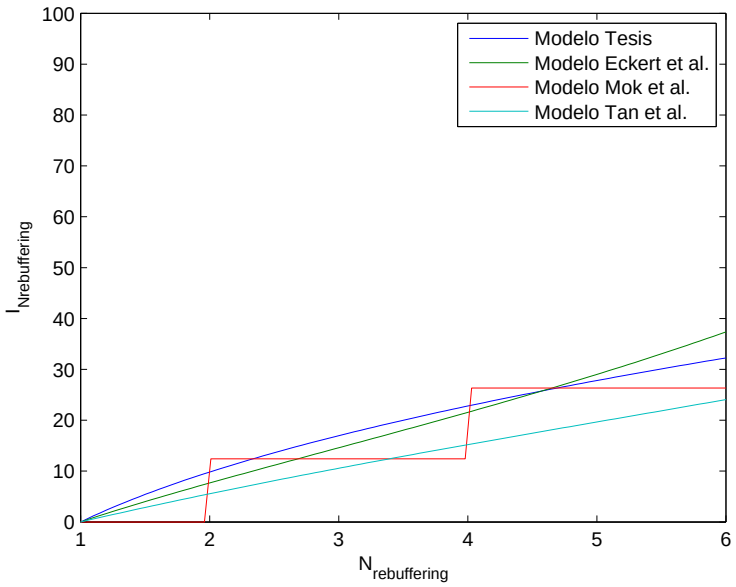


Figura 5.16: Efecto del número de eventos de rebuffering: comparativa con otros modelos

- Se seleccionaron tres niveles de calidad de vídeo, correspondientes a valores de MOS de 5, 3 y 1.
- El número máximo de conmutaciones estudiadas fue de 5.

Teniendo esto en cuenta, se definieron las siguientes trayectorias de adaptación, cada una de las cuales se aplica a dos secuencias de vídeo en el experimento, haciendo un total de 20 secuencias:

- $Q_5\,Q_3$
- $Q_3\,Q_1$
- $Q_5\,Q_1$
- $Q_5\,Q_3\,Q_5\,Q_3$
- $Q_5\,Q_3\,Q_1$
- $Q_5\,Q_3\,Q_5\,Q_1$
- $Q_5\,Q_1\,Q_3\,Q_1$
- $Q_3\,Q_1\,Q_3\,Q_1$
- $Q_3\,Q_1\,Q_3\,Q_5\,Q_3\,Q_1$

- $Q_5\,Q_1\,Q_3\,Q_1\,Q_3\,Q_1$

Para cada trayectoria de adaptación se ha obtenido un valor de calidad "total" agregando las valoraciones de los usuarios. Sin embargo, el objetivo del experimento es obtener información sobre la degradación adicional (en caso de que exista) asociada a la conmutación entre calidades. Para ello, el procedimiento que se ha aplicado es el siguiente:

$$I_{\Delta Q} = Q_{teórica} - Q_{experimento} = \frac{1}{T}\sum_i t_i \cdot Q_i - Q_{experimento} \tag{5.14}$$

Como se puede ver en la ecuación 5.14, el valor de degradación se obtiene sustrayendo al valor de calidad "teórica" (que es la media ponderada de los niveles de calidad usando el tiempo de reproducción de cada nivel de calidad como peso) el valor de calidad obtenido en el experimento.

En la tabla 5.5 se recogen los resultados obtenidos en el experimento, en escala R. Como se puede ver, el rango de valores de $I_{\Delta Q}$ es bastante amplio. Para ciertos experimentos, como por ejemplo $Q_3\,Q_1$, apenas se ha registrado degradación adicional asociada al cambio de calidad. En cambio, en otros experimentos, como por ejemplo $Q_5\,Q_1$, la degradación registrada ha sido muy importante. Sorprende también que hay varios experimentos, como $Q_5\,Q_1\,Q_3\,Q_1$ o $Q_3\,Q_1\,Q_3\,Q_5\,Q_3\,Q_1$, donde la calidad no se degrada, sino que mejora con respecto a la calidad teórica (valores de degradación negativos no despreciables).

Tabla 5.5: Resultados del experimento de evaluación de calidad en escenarios de adaptación del nivel de calidad

Trayectoria de adaptación	$I_{\Delta Q}$ (en escala R)
$Q_5\,Q_3$	-1,872
$Q_3\,Q_1$	-0,398
$Q_5\,Q_1$	39,91
$Q_5\,Q_3\,Q_5\,Q_3$	-2,485
$Q_5\,Q_3\,Q_1$	16,12
$Q_5\,Q_3\,Q_5\,Q_1$	15,73
$Q_5\,Q_1\,Q_3\,Q_1$	-18,68
$Q_3\,Q_1\,Q_3\,Q_1$	-2,244
$Q_3\,Q_1\,Q_3\,Q_5\,Q_3\,Q_1$	-15,10
$Q_5\,Q_1\,Q_3\,Q_1\,Q_3\,Q_1$	-8,976

A la vista de los resultados obtenidos en el experimento, se puede diseñar un algoritmo que modele dichos resultados, ver algoritmo 1.

Analizando los resultados se puede plantear la idea de que la degradación o mejora que se produce en la calidad percibida como consecuencia de la adaptación de la calidad de vídeo está fuertemente condicionada por el nivel de calidad más "atípico"

Entrada: Conjunto Q de calidades que forman la trayectoria de adaptación, $Q = [Q_1...Q_n]$
Salida: Degradación asociada a la adaptación de calidad de vídeo, $I_{\Delta Q}$
$\Delta_Q \leftarrow Q_{max} - Q_{min}$;
if $\Delta_Q \leq 50$ **then**
  $I_{\Delta Q} \leftarrow 0$;
else
  $Q_m \leftarrow$ media aritmética de los valores de Q;
  $d_{max} \leftarrow \infty$;
  $Q_o \leftarrow null$;
  foreach $Q_i \in Q$ **do**
    $d \leftarrow |Q_i - Q_m|$;
    if $d < d_{max}$ *or* $(d = d_{max}$ *and* $Q_i < Q_o)$ **then**
      $d_{max} \leftarrow d$;
      $Q_o \leftarrow Q_i$;
    end
  end
  if $Q_o > Q_m$ **then**
    $I_{\Delta Q} \leftarrow \alpha < 0$; `// mejora en la calidad`
  else
    $I_{\Delta Q} \leftarrow \alpha > 0$; `// degradación en la calidad`
  end
end

Algoritmo 1: Algoritmo de estimación de la degradación asociada a la adaptación de calidad de vídeo

que se ha reproducido. Por ejemplo, en trayectorias de adaptación donde predominan niveles de calidad altos, cuando se conmuta a un nivel de calidad bajo, se produce degradación adicional (trayectoria $Q_5\,Q_3\,Q_5\,Q_1$). El caso contrario también aplica. En trayectorias donde los niveles de calidad suelen ser bajos, el que se conmute a un nivel de calidad superior se recompensa en las valoraciones de calidad de los usuarios (trayectoria $Q_5\,Q_1\,Q_3\,Q_1$). Por otro lado, si los niveles de calidad entre los que se conmuta no están demasiado alejados entre sí, no se registra degradación ni mejora adicional en las valoraciones de calidad de los usuarios.

Así pues, el algoritmo propuesto se basa en buscar el nivel de calidad más alejado de la calidad media. Dependiendo de si este valor es mayor o menor que la media, ésto supondrá un incremento (mejora) o un decremento (degradación) de la calidad. Más concretamente, como se desprende del pseudocódigo propuesto, si el rango de calidad de la trayectoria de adaptación (en escala R) es menor de 50 (menor que 2 en escala MOS), no hay variación significativa de calidad entre la calidad teórica y la calidad real, por lo que $I_{\Delta Q} = 0$. Si la diferencia es mayor, se busca el valor de calidad Q_o que esté más alejado del valor de calidad medio Q_m. Si el nivel de calidad Q_o es mayor que Q_m entonces se obtiene una mejora en la calidad, es decir $I_{\Delta Q} < 0$. Si por el contrario,

Q_o es menor que Q_m se produce una degradación en la calidad, $I_{\Delta Q} > 0$.

Otro aspecto destacable de los resultados obtenidos ponen de manifiesto un cierto "efecto memoria" en las valoraciones de los usuarios. Más concretamente, si se examinan los resultados obtenidos en las dos últimas trayectorias de adaptación ($Q_3\,Q_1\,Q_3\,Q_5\,Q_3\,Q_1$ y $Q_5\,Q_1\,Q_3\,Q_1\,Q_3\,Q_1$) se puede ver que la mejora de la calidad es mayor en la primera trayectoria. Esto puede ser explicado al observar que el valor de calidad atípico (Q_5 en ambos casos) acontece en instantes de tiempo diferentes en ambas trayectorias de adaptación. En $Q_3\,Q_1\,Q_3\,Q_5\,Q_3\,Q_1$ el valor de Q_5 aparece en la última parte de la secuencia, mientras que en $Q_5\,Q_1\,Q_3\,Q_1\,Q_3\,Q_1$ aparece al principio de la misma. Al aparecer al principio, los usuarios pueden haber "olvidado" que hubo un fragmento de vídeo de alta calidad, mientras que al aparecer al final dicho valor de calidad toma más relevancia.

La diversidad de resultados que se han obtenido en este experimento hacen difícil la propuesto de un modelo cuantitativo de la degradación o mejora de calidad asociada a la adaptación de calidad. Por ello, en el algoritmo propuesto no se han definido valores concretos de degradación y de mejora de calidad. Además, la detección del efecto memoria abre un abanico de posibilidades bastante amplio, lo cual lleva a plantear una línea de investigación futura que se centre en el análisis del efecto de los mecanismos de adaptación en la calidad percibida. Esta línea de investigación deberá plantear nuevos experimentos de evaluación subjetiva de calidad que permitan obtener resultados más concluyentes que los obtenidos hasta el momento.

5.4. Análisis de la influencia de la red en las variables del modelo

El modelo propuesto en la sección anterior utiliza una serie de variables (tiempo de rebuffering, número de eventos de rebuffering, etc.), cuyo origen está íntimamente ligado con el rendimiento de la red, con una serie de parámetros de la implementación del cliente de vídeo del usuario (tamaño del buffer de vídeo, tipo de algoritmo de adaptación, etc.) y con el nivel de calidad seleccionado en cada momento por dicho cliente, el cual depende a su vez del algoritmo de adaptación que se esté aplicando.

Como se ha comentado anteriormente, la pila de protocolos que se utiliza típicamente para desplegar servicios de vídeo OTT está formada por HTTP (+ MPEG-DASH) <->TCP <->IP <->L2 <->L1.

La capa de transporte TCP es la que hace que la entrega de paquetes se lleve a cabo de manera ordenada y sin errores, a cambio de un cierto retardo (consecuencia, entre otras cosas de retransmisiones, efectos de las ventanas de congestión, etc.). TCP implementa varios mecanismos de control: control de errores, control de flujo y control de congestión. El primero se encarga de que la información se entregue sin errores

mientras que el segundo tiene como objetivo evitar que el emisor sature al receptor. El control de congestión intenta favorecer el rendimiento de la red evitando que ésta se colapse debido a una carga de tráfico demasiado elevada y juega un papel fundamental a la hora de estudiar el rendimiento que pueden alcanzar los servicios de streaming de vídeo sobre TCP.

Además, el nivel de calidad que el cliente solicite en cada momento influirá en el tamaño de los fragmentos de vídeo que se deben transmitir por la red y en la tasa de bit que el decodificador de vídeo del cliente espera. Por otro lado, se debe tener en cuenta que las implementaciones de clientes de vídeo suelen contar con un buffer con el que amortiguar los efectos de la red.

Así pues, el objetivo de esta sección es estudiar la relación entre los parámetros de la red, el nivel de calidad de vídeo que se está transmitiendo, el tamaño del buffer del cliente y la implementación del algoritmo de adaptación de calidad.

5.4.1. Aproximación al problema de manera analítica

Si se conoce la expresión del goodput de la capa TCP y la tasa de codificación del vídeo en función del tiempo, entonces es posible establecer una función analítica que permita calcular el nivel de ocupación del buffer de un cliente de vídeo en función del tiempo.

Más concretamente, se puede definir $B(t,p)$ como los segundos de vídeo que están almacenados en el buffer del cliente en función del tiempo, t, para un valor de probabilidad de pérdidas de paquete, p de acuerdo a la ecuación 5.15.

$$B(t,p) = B_0 + \int_{t_0}^{t} \frac{\beta(t,p)}{\lambda(t)} - r(t)\,\mathrm{d}x \tag{5.15}$$

En esta ecuación B_0 representa el nivel inicial del buffer (típicamente $B_0 = 0$), $\beta(t,p)$ es la tasa de bit útiles recibidos (goodput TCP), $\lambda(t)$ es el bitrate del vídeo recibido y $r(t)$ es una función tal que $r(t) = 1$ si la reproducción del vídeo está en curso y $r(t) = 0$ si la reproducción está detenida (como consecuencia del buffering inicial o de un evento de rebuffering).

Como se puede ver, la ecuación 5.15 modela dos procesos que interactúan con el buffer: el primer proceso (recepción de vídeo) aumenta la ocupación del buffer a una tasa $\frac{\beta(t,p)}{\lambda(t)}$, mientras que el segundo proceso (reproducción del vídeo) descarga el buffer a una tasa de 1 mientras se reproduce vídeo o a una tasa de 0 mientras la reproducción está detenida.

Así pues, si todos los factores de esta expresión fuesen conocidos, partir de ella se podrían calcular los parámetros necesarios para el modelo de calidad percibida, ya que de el nivel del buffer del cliente depende que se produzcan eventos de rebuffering.

Sin embargo, como se explica a continuación, realizar este estudio de manera analítica conlleva ciertos problemas.

Uno de los modelos analíticos de TCP más conocido es el de [Padhye et al., 2000]. En dicho trabajo se pueden encontrar expresiones que modelan la tasa de envío de un emisor que utiliza TCP Reno, en función del RTT, el tamaño máximo de la ventana de congestión, la tasa de pérdidas y el timeout de retransmisión. Otra expresión que se puede encontrar en dicho trabajo modela la tasa de bit que el receptor percibe.

Sin embargo, es importante destacar una serie de consideraciones en cuanto al modelo de [Padhye et al., 2000]. La expresión del goodput que ofrece es una expresión que modela la tasa de paquetes que recibe el cliente por unidad de tiempo en régimen permanente, ya que dicho modelo está planteado para modelar una descarga de un fichero en la que el servidor tiene infinitos datos que enviar. Dependiendo de la implementación del cliente de streaming adaptativo, pueden darse casos donde la expresión del goodput de [Padhye et al., 2000] no modele adecuadamente el streaming de vídeo. Por ejemplo, si el cliente no anticipa las peticiones de nuevos fragmentos de vídeo a la finalización de la recepción del fragmento que se está transmitiendo, habrá intervalos de tiempo (un RTT) donde el cliente no esté recibiendo paquetes de vídeo.

Por otro lado, el modelo de [Padhye et al., 2000] proporciona una expresión aproximada de la ventana de congestión que puede llegar a obtener un flujo TCP. Sin embargo, las suposiciones en cuanto a independencia estadística que se realizan en este modelo pueden llevar a errores cuando se trata de varios flujos compitiendo entre sí. Por ejemplo, la tasa de pérdidas, en general no es independiente de los valores de los parámetros AIMD (Additive Increase/Multiplicative Decrease) de los flujos que compiten por la red [Shorten et al., 2006], ni tampoco el RTT, ya que éste dependerá del nivel de ocupación de las colas de los routers que tienen que atravesar los paquetes entre el origen y el destino de la comunicación.

Estudios más recientes se apoyan en enfoques basados en teoría de fluidos y en disciplinas de colas activas. Sin embargo, el enfoque de fluidos presenta ciertas dificultades a la hora de modelar colas de tipo drop tail.

Además, los modelos analíticos que estudian el control de congestión de TCP suelen estar orientados a escenarios donde todas las fuentes de tráfico compiten continuamente por el ancho de banda. Esta suposición no es directamente aplicable al caso del streaming de vídeo adaptativo, donde en condiciones normales, cuando el buffer del cliente esté lleno no se seguirán solicitando fragmentos de vídeo hasta que haya espacio disponible en el buffer.

En todo caso, en escenarios de streaming de vídeo adaptativo, se podrían aplicar los modelos analíticos de TCP al régimen permanente de la transmisión, es decir, cuando los algoritmos de adaptación de todos los clientes han convergido a un nivel de calidad

estable (suponiendo que el tráfico de fondo de la red permitiese esta convergencia) y están simultáneamente descargando fragmentos de vídeo. Sin embargo, uno de los objetivos del modelo planteado en la tesis es cuantificar el efecto que supone a la calidad percibida las degradaciones que se producen en el régimen transitorio, es decir, cuando el nivel de calidad no es el adecuado, cuando el tráfico de fondo obliga a realizar un cambio de nivel de calidad o cuando la red no es capaz de soportar el tráfico de vídeo que se desea cursar, dando lugar a eventos de rebuffering.

Todas estas razones han llevado a abordar este problema mediante herramientas de simulación de red.

5.4.2. Aproximación al problema mediante simulación de red

En esta sección se aborda el análisis de la influencia de la red en las variables del modelo de degradación de calidad (tiempo de buffering inicial, tiempo de rebuffering y número de eventos de rebuffering) mediante herramientas de simulación de red.

En primer lugar, se ha llevado a cabo una revisión de las herramientas de simulación de red más importantes y más utilizadas actualmente. Dicha revisión se puede encontrar en el apéndice D. Como resultado de esta revisión se ha seleccionado OMNeT++ como herramienta de simulación, por las siguientes razones:

- El carácter modular y extensible de OMNeT++ y del framework INET es la característica que más peso ha tenido a la hora de seleccionar OMNeT++ como herramienta a utilizar.
- Interfaz gráfica con funcionalidades de generación de gráficas, estadísticas, animaciones, etc.
- La cantidad y calidad de la documentación es aceptable.
- Herramienta gratuita para uso académico, con licencia similar a GNU-GPL .

5.4.2.1. Modelo de simulación de streaming de vídeo adaptativo sobre TCP en OMNeT++

En ninguna de las herramientas de simulación analizadas existen módulos con los que simular las características propias del streaming adaptativo sobre TCP, por lo que ha sido necesario desarrollar un nuevo modelo, a partir de los que ya están desarrollados en la herramienta.

En este contexto el framework INET (que implementa la torre de protocolos TCP/IP sobre OMNeT++) destaca por la claridad de su diseño y por una serie de clases e interfaces bien definidas sobre las que poder implementar nuevos modelos. El framework

INET es la pieza básica de simulación de redes y protocolos de comunicación en OMNeT++. INET proporciona implementaciones de protocolos como IPv4, IPv6, TCP, SCTP o UDP, además de diversos modelos de aplicación. Ofrece también modelos MPLS con señalización RSVP-TE y LDP. En la capa de enlace proporciona modelos PPP, Ethernet y 802.11. En cuanto al encaminamiento, éste puede ser configurado específicamente para cada escenario de simulación o bien se pueden utilizar implementaciones concretas de protocolos de encaminamiento. Para los intereses de esta tesis son de especial importancia los niveles de transporte y de aplicación dentro de la torre de protocolos TCP/IP.

La interfaz entre el protocolo TCP y el nivel de aplicación se puede realizar mediante la utilización de la clase TCPSocket. Esta clase facilita la gestión de las conexiones TCP desde los módulos de aplicación mediante sus métodos bind(), listen(), connection(), etc.

A continuación, en la figura 5.17 se muestra un diagrama de clases que representa el diseño elegido para el modelo de streaming de vídeo adaptativo.

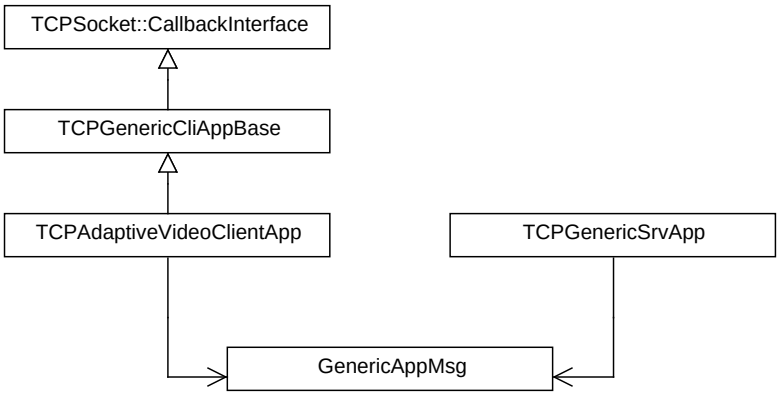


Figura 5.17: Diagrama de clases del modelo de simulación de streaming de vídeo adaptativo

La lógica del streaming adaptativo está recogida en la clase TCPAdaptiveVideoClientApp, que a su vez extiende a TCPGenericCliAppBase, clase proporcionada por INET como base para el desarrollo del lado de cliente en aplicaciones cliente-servidor. El cliente lleva a cabo dos procesos ligados a un buffer de recepción: el primero de ellos se encarga de ir llenando el buffer con fragmentos de vídeo de un nivel de calidad determinado y el segundo se encarga de ir consumiendo esos fragmentos de vídeo. La clase cliente se comunica con un servidor implementado en la clase TCPGenericSrvApp. Este módulo acepta conexiones TCP y espera recibir mensajes de clase GenericAppMsg. Este tipo de mensaje es especialmente útil para la simulación del streaming adaptativo

ya que contiene un campo en el que el cliente puede indicar al servidor el tamaño de la respuesta que espera.

El algoritmo de adaptación que se ha implementado inicialmente es el siguiente:

1. Cuando se lanza la simulación el buffer está vacío y se comienza el llenado del mismo utilizando el nivel de calidad más bajo.
2. Cuando el buffer llega a cierta capacidad (configurable) comienza la reproducción, eliminando del buffer un elemento por segundo (se asume que cada paquete que se solicita al servidor es de 1 segundo de vídeo).
3. Reglas de adaptación
   a) Si el buffer llega a su capacidad máxima, se aumenta el nivel de calidad solicitado.
   b) Si el buffer baja de cierta capacidad (configurable), se disminuye el nivel de calidad solicitado.

OMNeT++ proporciona potentes mecanismos para la generación de estadísticas y logs personalizados. Estos mecanismos se han aprovechado para generar una serie de trazas y gráficas de especial interés en el caso del streaming adaptativo:

- Nivel del buffer en función del tiempo: permite analizar el comportamiento de los procesos de llenado y vaciado del buffer.
- Control de reproducción del vídeo.
  - PlaybackPointer: variable que representa qué instante de vídeo se está reproduciendo en cada momento.
  - PlaybackStatus: variable booleana que indica si se está reproduciendo vídeo o se está en un estado de rebuffering.
- Nivel de calidad solicitado en cada petición que se realiza al servidor.

Como ejemplo, en la figura 5.18 se muestran las trazas que genera una simulación de streaming de vídeo adaptativo:

5.4.2.2. Objetivos de la simulación y diseño del escenario

En esta sección se describen los objetivos que se persiguen con las simulaciones realizadas y el proceso de diseño del escenario de simulación que se ha planteado para estudiar el efecto de la red y de la implementación del cliente de vídeo sobre la calidad del streaming de vídeo adaptativo.

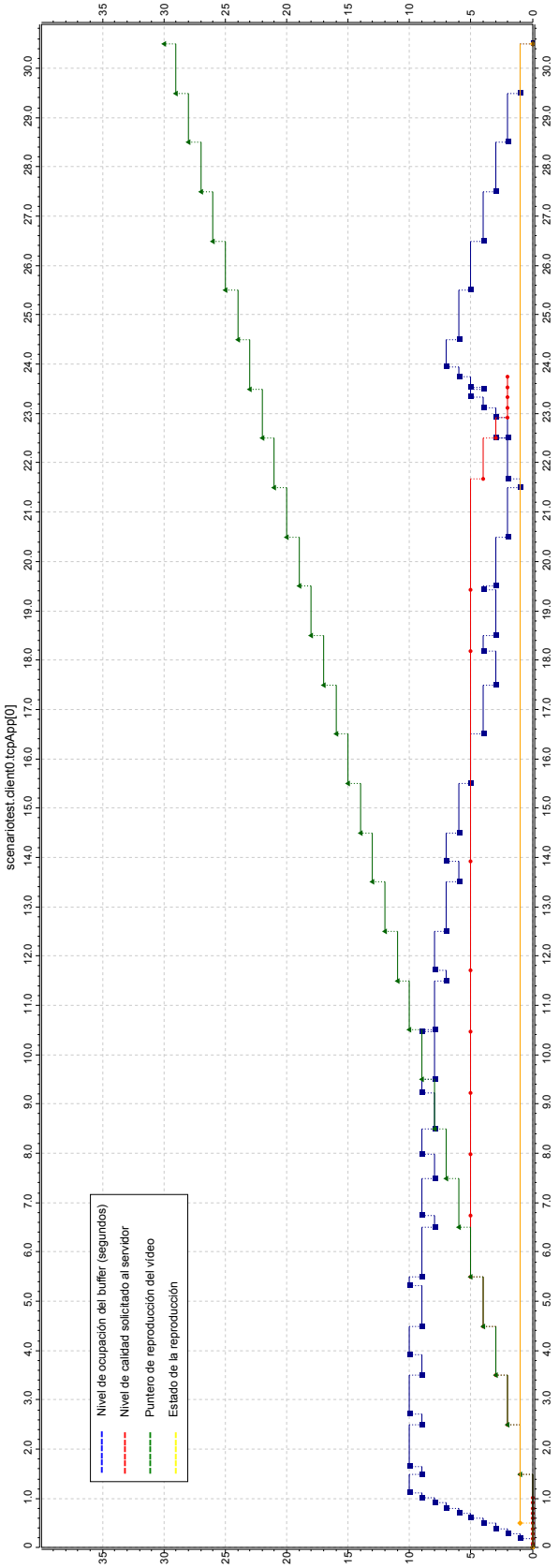


Figura 5.18: Ejemplo de trazas de simulación de streaming de vídeo adaptativo

Objetivos El principal objetivo de las simulaciones que se van a realizar es estudiar el efecto que las condiciones de la red y los parámetros de implementación de los clientes tienen en la calidad percibida por los usuarios del servicio de streaming de vídeo adaptativo sobre TCP. Para ello, en cada simulación se van a recoger las siguientes variables:

- Tiempo de buffering inicial
- Tiempo total de rebuffering
- Número de eventos de rebuffering

En las simulaciones realizadas se estudia el efecto que tiene la capacidad de los enlaces en las variables de calidad de experiencia, suponiendo un escenario de hora cargada, donde todos los usuarios del servicio hacen uso del mismo en una franja de tiempo determinada

Diseño del escenario de simulación La topología de la red que se ha seleccionado es una topología en árbol, donde la raíz está formada por el servidor de streaming (encargado de enviar los fragmentos de vídeo con la calidad que soliciten los clientes) y un servidor web (encargado de atender las peticiones del tráfico HTTP que se utilizará como tráfico de fondo), los nodos de los niveles intermedios son routers y las hojas son los clientes finales (ver figura 5.19).

En cuanto a la capacidad de los enlaces de esta topología, se han realizado las siguientes consideraciones:

Canales D_i: en el informe anual de 2012 de la Comisión Nacional de los Mercados y la Competencia, en el apartado de "líneas de banda ancha fijas por segmento y velocidad", se presentan los siguientes datos para el sector residencial [CNMC, 2012], (tabla 5.6):

Tabla 5.6: Líneas de banda ancha fijas por segmento y velocidad [CNMC, 2012]

Velocidad (v)	**Número de líneas**	**Porcentaje**
$v < 2\,\text{Mbps}$	196.435	2,14 %
$2\,\text{Mbps} \leq v < 10\,\text{Mbps}$	2.984.077	32,43 %
$10\,\text{Mbps} \leq v < 30\,\text{Mbps}$	5.017.669	54,54 %
$30\,\text{Mbps} \leq v < 50\,\text{Mbps}$	587.776	6,39 %
$50\,\text{Mbps} \leq v$	414.738	4,51 %

Teniendo en cuenta los datos de la tabla 5.6, la decisión que se ha tomado es utilizar en las simulaciones la misma distribución de velocidades de acceso que proporciona la CNMC, tomando como "representantes" de cada franja las siguientes velocidades (tabla 5.7):

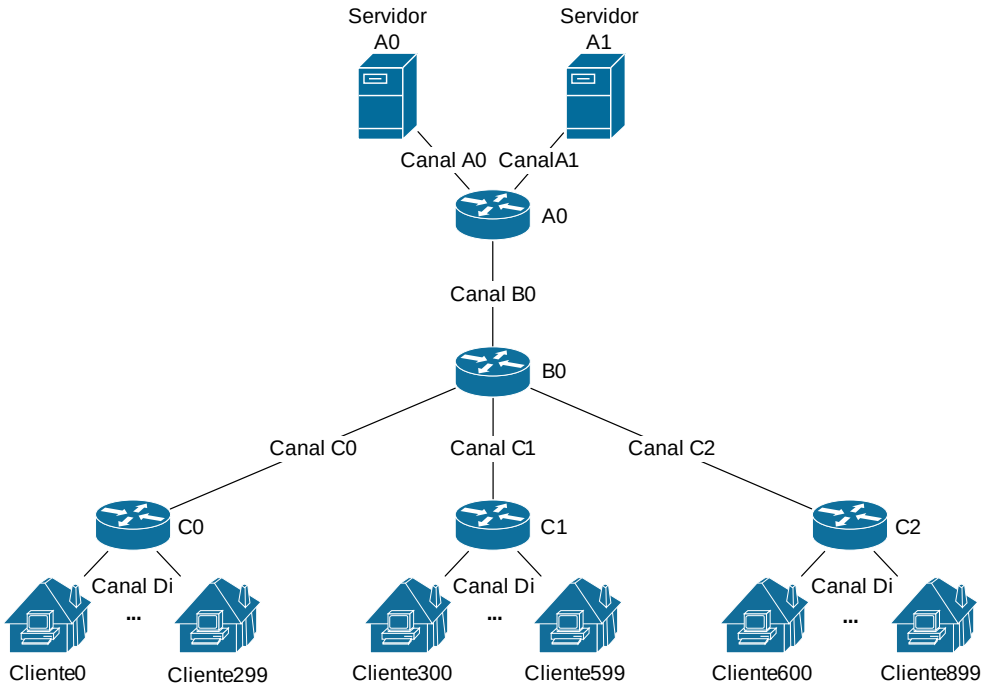


Figura 5.19: Topología de la red simulada

Tabla 5.7: Velocidades consideradas en los canales D_i

Velocidad	**Porcentaje**
2 Mbps	2 %
6 Mbps	32 %
20 Mbps	55 %
30 Mbps	6 %
100 Mbps	5 %

Enlaces router-router / router-servidor: la capacidad de estos enlaces se especificará en cada experimento con el objetivo de controlar el nivel de saturación que alcanzará la red.

A continuación se describen otros parámetros que se han utilizado en el conjunto de simulaciones realizadas:

Número de clientes por cada router C_i: como se puede ver en la figura 5.19, para la realización de las simulaciones se van a asignar 300 usuarios a cada router C_i.

Penetración del servicio: de acuerdo a los datos de informe de 2012 de la CNMC, el porcentaje de líneas de banda ancha con servicio de televisión contratado es del 23,4 % (con los accesos de cable aportando la mayoría de abonados de TV a la cifra total). En principio, en las simulaciones se asumirá que un 25 % de los usuarios harán uso del servicio de streaming de vídeo . Esta es una suposición muy ambiciosa, la cual supone

una tendencia alcista tanto en el uso de los servicios de vídeo como en la adopción del vídeo OTT.

Duración del vídeo simulado: la duración del vídeo que los usuarios de la simulación consumen es de 600 segundos.

Tráfico de fondo: además del tráfico de vídeo, se va a cursar tráfico HTTP para simular tráfico de fondo que compite por los recursos con el tráfico de vídeo. El tráfico de fondo se va a generar utilizando el framework HttpTools, disponible como una librería de INET. La configuración de este tráfico es la siguiente:

- Número de peticiones por sesión $\sim \mathcal{N}(20, 10)$
- Tiempo entre peticiones (s) $\sim \mathcal{N}(300, 60)$
- Tamaño de petición (bytes) $\sim \mathcal{N}(600, 100)$
- Tamaño página (bytes) $\sim Exp(2000)$
- Imágenes por página $\sim \mathcal{U}(0, 20)$
- Tamaño imagen (bytes) $\sim Exp(20000)$

5.4.2.3. Resultados de las simulaciones

Experimento 1 En el primer experimento de simulación se seleccionaron las capacidad de canal que se indican en la tabla 5.8.

Tabla 5.8: Capacidades de los canales para el experimento de simulación 1

Sim.	**Canal C_0**	**Canal C_1**	**Canal C_2**	**Canal B_0**	**Canal A_0**	**Canal A_1**
1	250 Mbps	250 Mbps	250 Mbps	600 Mbps	1 Gbps	1 Gbps
2	250 Mbps	250 Mbps	250 Mbps	675 Mbps	1 Gbps	1 Gbps
3	250 Mbps	250 Mbps	250 Mbps	750 Mbps	1 Gbps	1 Gbps
4	500 Mbps	500 Mbps	500 Mbps	2 Gbps	5 Gbps	5 Gbps
5	750 Mbps	750 Mbps	750 Mbps	4 Gbps	5 Gbps	5 Gbps
6	1 Gbps	1 Gbps	1 Gbps	5 Gbps	5 Gbps	5 Gbps

El conjunto de representaciones de vídeo (modeladas por su tasa de bit) que los usuarios pueden seleccionar es el siguiente: $Q = \{1; 1,5; 2; 4; 8; 12\} Mbps$.

Así pues, la capacidad (C) que requieren los canales C_i para transportar el tráfico de vídeo se presenta en la ecuación 5.16, donde N_i es el número de usuarios que tienen una cierta capacidad de canal de acceso D_i y $Q_{i\,max}$ es el nivel de calidad máximo que el canal D_i permite a cada usuario.

$$C = \sum_i N_i \cdot Q_{i\,max} \tag{5.16}$$

Evaluando la ecuación 5.16, se obtiene que la capacidad necesaria en los canales C_i para cursar la máxima demanda de tráfico de vídeo que los clientes pueden solicitar es de 697,5 Mbps.

Teniendo en cuenta este resultado y las capacidades de canal indicadas en la tabla 5.8, se puede ver que las simulaciones 1 y 2 representan escenarios con cuellos de botella tanto en los canales C_i como en el canal B_0. Las simulaciones 3 y 4 son escenarios con cuello de botella en los canales C_i. Las simulaciones 5 y 6 son escenarios donde los enlaces tienen capacidad suficiente para ofrecer vídeo a todos los clientes con la tasa de bit máxima que su canal de acceso les permite (sin contar el tráfico de fondo).

Además de las capacidades de los enlaces, es importante tener en cuenta los detalles de implementación utilizados en los clientes. El algoritmo de adaptación sigue las siguientes reglas:

- Conmutación a un nivel de calidad superior cuando el buffer se llena.
- Conmutación a un nivel de calidad inferior mientras el nivel del buffer esté en una zona crítica (ocupación del buffer menor de 4 segundos de vídeo). Es decir, mientras el nivel del buffer se encuentre en la zona crítica, en cada petición de fragmentos de vídeo, el nivel de calidad solicitado se reduce en una unidad.

Los parámetros utilizados en el algoritmo de adaptación son los siguientes:

- Tamaño del buffer de vídeo: 10 segundos
- Zona crítica: 4 segundos

Una vez descritas las particularidades del experimento de simulación, a continuación se presentan los resultados de cada una de las simulaciones.

En primer lugar se presentan los valores medios y las desviaciones para cada uno de los parámetros analizados (tabla 5.9).

Tabla 5.9: Resultados agregados del experimento de simulación 1

Simulación	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
1.1	1,047	1,081	31,932	18,309	8,236	3,849
1.2	1,045	1,111	26,439	14,213	7,418	3,199
1.3	1,042	1,207	18,798	9,485	6,027	2,326
1.4	0,634	0,660	6,564	5,793	2,844	1,931
1.5	0,480	0,416	0,922	1,499	0,507	0,689
1.6	0,448	0,371	0	0	0	0

Como se puede ver en la tabla 5.9, y de acuerdo al diseño de los experimentos comentado anteriormente, en el escenario 1.6 hay capacidad suficiente para transportar

tanto el vídeo como el tráfico de fondo. Sin embargo, en el escenario 1.5, el tráfico de fondo hace que el vídeo sufra cierta degradación.

En el resto de escenarios, los enlaces tienen una capacidad muy inferior a la requerida para poder soportar el tráfico de vídeo a la máxima calidad que permiten los enlaces D_i y el tráfico de fondo. Sin embargo, los canales C_i sí que tienen capacidad para soportar el tráfico de vídeo de todos los usuarios a los que dan servicio, si estos solicitaran niveles de calidad bajos, ya que $2\,Mbps \cdot 75\,usuarios = 150\,Mbps < 250\,Mbps$. En el caso del canal B_0, en el primer escenario de simulación, cuenta con una capacidad de 600 Mbps, la cual sería suficiente para soportar la demanda de todos los usuarios de vídeo si éstos solicitaran un nivel de calidad adecuado: $2\,Mbps \cdot (75 \cdot 3)\,usuarios = 450\,Mbps < 600\,Mbps$.

Así pues, el hecho de que se produzcan tiempos de rebuffering tan elevados puede indicar que el algoritmo de adaptación utilizado no es capaz de adaptarse a las condiciones de la red de manera eficiente.

Además de los resultados agregados mostrados anteriormente, es interesante también analizar el nivel de degradación que se obtiene agrupando a los usuarios por la velocidad de su canal de acceso (canales D_i).

Tabla 5.10: Resultados de la simulación 1.1

Canal D_i	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
2 Mbps	4,008	2,634	2,875	4,980	1	1,732
6 Mbps	1,302	0,925	13,551	7,631	4,486	2,233
20 Mbps	0,865	1,002	41,920	14,676	10,260	2,958
30 Mbps	0,803	0,966	41,870	14,078	10,200	2,731
100 Mbps	0,840	1,164	34,670	10,246	9,333	1,923

Tabla 5.11: Resultados de la simulación 1.2

Canal D_i	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
2 Mbps	3,304	0,706	4,814	5,059	1,333	1,527
6 Mbps	1,336	0,877	12,678	6,699	4,375	1,788
20 Mbps	0,890	1,167	32,984	11,480	8,854	2,604
30 Mbps	0,462	0,900	40,427	10,076	10,400	1,454
100 Mbps	1,051	1,121	29,837	10,139	8,750	1,289

Al analizar estos resultados, se pone de manifiesto un comportamiento inesperado: en determinados experimentos, los usuarios con canales de acceso de mayor capacidad consiguen peor rendimiento que aquellos con conexiones más limitadas. Este comportamiento se puede ver claramente en las simulaciones 1.3, 1.4 y 1.5, reflejado especialmente en las variables “Tiempo de rebuffering” y “Número de eventos de rebuffering”.

Tabla 5.12: Resultados de la simulación 1.3

Canal D_i	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
2 Mbps	2,498	0,011	7,971	6,266	2,33	1,528
6 Mbps	1,507	1,446	10,921	5,785	3,958	1,614
20 Mbps	0,799	0,994	22,312	8,121	6,935	1,885
30 Mbps	0,546	0,965	22,723	8,739	7,533	1,356
100 Mbps	0,996	1,023	27,832	9,681	8,167	1,467

Tabla 5.13: Resultados de la simulación 1.4

Canal D_i	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
2 Mbps	2,485	0,0004	0	0	0	0
6 Mbps	1,050	0,559	2,3	2,144	1,083	0,931
20 Mbps	0,443	0,579	7,385	4,078	3,423	1,385
30 Mbps	0,233	0,258	9,876	6,370	4,600	2,098
100 Mbps	0,138	0,290	21,234	5,443	6	1,477

Tabla 5.14: Resultados de la simulación 1.5

Canal D_i	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
2 Mbps	2,484	0,0004	0	0	0	0
6 Mbps	0,842	0,120	0,895	1,434	0,431	0,624
20 Mbps	0,284	0,242	0,979	1,593	0,545	0,704
30 Mbps	0,233	0,26	0,224	0,686	0,2	0,561
100 Mbps	0,138	0,301	1,610	1,511	1,083	0,793

Tabla 5.15: Resultados de la simulación 1.6

Canal D_i	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
2 Mbps	2,484	0,000125	0	0	0	0
6 Mbps	0,828	0,000211	0	0	0	0
20 Mbps	0,249	0,000159	0	0	0	0
30 Mbps	0,166	0,000061	0	0	0	0
100 Mbps	0,05	0,000274	0	0	0	0

La causa de este comportamiento es el algoritmo de adaptación utilizado en el modelo de simulación.

Como se comentó anteriormente, se está utilizando un algoritmo simplificado en el que el cambio a una representación de mayor calidad se realiza cuando el buffer se llena. El principal problema que presenta esta estrategia de adaptación es que el cambio a un nivel de calidad mayor se realiza sin tener en cuenta si la red va a ser capaz de soportar el nuevo nivel de demanda de tráfico. Por ejemplo, si se toma un cliente de alguna de las simulaciones anteriores, el cual está solicitando vídeo codificado a 4 Mbps y en ese momento la tasa de bit máxima que la red permite es de 6 Mbps, el buffer de vídeo del cliente se terminará llenando y en ese momento el cliente conmutará al siguiente nivel de calidad, que en este caso es de 8 Mbps. Al conmutar, la red no será capaz de proporcionar los fragmentos de vídeo codificados a 8 Mbps a tiempo, por lo que el buffer se irá vaciando, produciéndose finalmente un evento de rebuffering y una nueva conmutación a un nivel de calidad inferior.

En la figura 5.20 se muestra una traza de la simulación que representa dicho comportamiento:

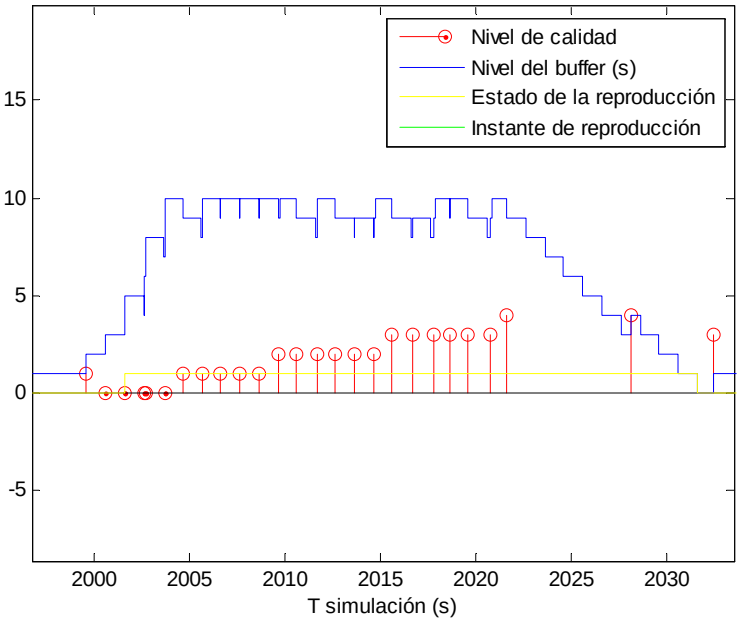


Figura 5.20: Comportamiento del algoritmo de adaptación simplificado

Como se puede ver en la traza de la figura, antes de realizar la conmutación al nivel máximo de calidad (en torno a $T_{simulacion} = 2020$), la transmisión era fluida y el buffer estaba en niveles altos. Esto significa que la red era capaz de soportar la demanda de

tráfico que dicho nivel de calidad imponía. Sin embargo, al pasar al siguiente nivel de calidad, el incremento en la demanda es demasiado alto y la red no es capaz de entregar a tiempo los fragmentos de vídeo, haciendo que el nivel del buffer baje rápidamente.

Experimento 2 Para mitigar este efecto, la solución adoptada ha sido modificar el algoritmo de adaptación, haciendo que la conmutación a un nivel superior solo se produzca si se estima que la tasa de bit disponible en la red es suficiente para soportar dicho nivel de calidad. Esta estimación se realiza registrando los tiempos de transmisión de los últimos paquetes solicitados.

Más concretamente, la estimación de la tasa de bit disponible se realiza mediante la ecuación 5.17, donde L_i representa la longitud del segmento de vídeo (en bits) ,T_j el tiempo (en segundos) necesario para la recepción de dicho segmento y n el número de fragmentos de vídeo utilizados para realizar la estimación.

$$\text{Tasa de bit estimada} = \frac{\sum_{i=1}^{n} L_i}{\sum_{j=1}^{n} T_j} \tag{5.17}$$

El parámetro n juega un papel importante en la estimación de la tasa de bit, ya que define la ventana temporal utilizada para realizar la estimación. En las simulaciones realizadas, el tamaño de la ventana de estimación utilizada ha sido de 5 fragmentos de vídeo con el objetivo de suavizar posibles valores excepcionales que se produzcan en la transmisión de algunos fragmentos de vídeo.

Así pues, se mantienen los parámetros del experimento 1, con las siguientes modificaciones en las reglas del algoritmo de adaptación:

- Conmutación a un nivel de calidad superior cuando el buffer se llena y tras comprobar que la tasa de bit del nivel de calidad al que se desea conmutar es menor que la que ofrece la red (en base a una estimación realizada por el cliente).
- Conmutación a un nivel de calidad inferior mientras el nivel del buffer esté en una zona crítica (ocupación del buffer menor de 4 segundos de vídeo). Mientras el nivel del buffer se encuentre en la zona crítica, en cada petición de fragmentos de vídeo, el nivel de calidad solicitado se reduce en una unidad.

Los resultados agregados de este segundo experimento se muestran en la tabla 5.16.

Si se comparan estos valores con los del experimento anterior se puede ver que se mejoran los resultados en todos los experimentos en un factor de aproximadamente el 30 %.

Sin embargo, al realizar el análisis agrupando a los usuarios por la velocidad de su canal de acceso, aunque en menor medida, se sigue produciendo el efecto comentado anteriormente.

Tabla 5.16: Resultados agregados del experimento de simulación 2

Simulación	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
2.1	1,010	1,245	20,937	12,742	6,070	2,970
2.2	1,004	1,252	14,431	9,381	4,640	2,503
2.3	0,933	1,054	10,281	6,365	3,770	1,887
2.4	0,556	0,568	4,107	4,629	1,813	1,752
2.5	0,469	0,406	0,489	1,110	0,306	0,588
2.6	0,450	0,375	0,000	0,000	0,000	0,000

Tabla 5.17: Resultados de la simulación 2.1

Canal D_i	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
2 Mbps	3,230	0,597	2,209	1,524	0,667	0,577
6 Mbps	1,381	1,232	7,959	5,262	2,983	1,603
20 Mbps	0,794	1,153	28,604	9,784	7,862	2,010
30 Mbps	0,747	1,234	25,634	9,317	7,217	1,872
100 Mbps	0,648	0,915	19,030	7,111	6,146	2,058

Tabla 5.18: Resultados de la simulación 2.2

Canal D_i	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
2 Mbps	2,998	0,716	1,168	1,147	0,417	0,433
6 Mbps	1,353	1,066	5,775	4,370	2,288	1,461
20 Mbps	0,834	1,292	19,418	7,984	5,945	1,951
30 Mbps	0,600	0,916	18,537	7,410	5,950	1,951
100 Mbps	0,656	1,119	13,432	6,249	4,792	1,812

Tabla 5.19: Resultados de la simulación 2.3

Canal D_i	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
2 Mbps	3,225	0,920	2,023	2,772	0,750	0,953
6 Mbps	1,222	0,753	5,816	3,815	2,413	1,288
20 Mbps	0,763	1,043	12,655	5,996	4,463	1,666
30 Mbps	0,603	1,034	12,723	6,678	4,667	1,939
100 Mbps	0,534	1,210	11,762	6,422	4,438	1,811

Tabla 5.20: Resultados de la simulación 2.4

Canal D_i	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
2 Mbps	2,485	0,001	0,000	0,000	0,000	0,000
6 Mbps	0,917	0,354	0,143	0,575	0,069	0,255
20 Mbps	0,368	0,487	5,255	3,636	2,439	1,329
30 Mbps	0,336	0,464	7,707	4,355	3,133	1,471
100 Mbps	0,114	0,220	12,653	6,276	4,667	1,631

Tabla 5.21: Resultados de la simulación 2.5

Canal D_i	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
2 Mbps	2,484	0,000	0,000	0,000	0,000	0,000
6 Mbps	0,851	0,132	0,000	0,000	0,000	0,000
20 Mbps	0,265	0,114	0,690	1,213	0,437	0,641
30 Mbps	0,184	0,069	0,217	0,507	0,150	0,339
100 Mbps	0,071	0,073	1,441	1,726	1,021	0,885

Tabla 5.22: Resultados de la simulación 2.6

Canal D_i	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
2 Mbps	2,484	0,000	0,000	0,000	0,000	0,000
6 Mbps	0,832	0,033	0,000	0,000	0,000	0,000
20 Mbps	0,251	0,023	0,000	0,000	0,000	0,000
30 Mbps	0,166	0,000	0,000	0,000	0,000	0,000
100 Mbps	0,050	0,000	0,000	0,000	0,000	0,000

Como se puede ver, los usuarios que acceden a la red a través de canales D_i de mayor capacidad obtienen mayores tiempos de rebuffering que el resto.

Experimento 3 Para continuar con el estudio se plantea un conjunto de cambios y mejoras de los parámetros de implementación del cliente:

- Aumento del tamaño del buffer: ayuda a contrarrestar las degradaciones en el rendimiento ofrecido por la red, ya que se incrementa el tiempo disponible para que el cliente se adapte a las condiciones de red antes de que el buffer quede vacío y se produzca un evento de rebuffering.

- Establecimiento de la zona crítica del buffer en 20 segundos: esta modificación permite que el algoritmo reaccione rápidamente a degradaciones en el rendimiento de la red.

- Aplicación de un factor reductor a la tasa de bit estimada, con el objetivo de que el algoritmo sea más conservador a la hora de conmutar a un nivel de calidad superior. Esta técnica se basa en los resultados proporcionados por [Wang et al., 2008], los cuales sugieren que se alcanza un rendimiento adecuado en el streaming sobre TCP cuando el throughput de la red es aproximadamente el doble de la tasa de bit de codificación de los flujos multimedia.

- Incremento del catálogo de representaciones del vídeo: el objetivo de esta modificación es que se reduzca la diferencia de tasa de bit entre los niveles de calidad

más altos, lo cual permite que el aumento de la demanda al conmutar a niveles de calidad superiores sea más escalonado. Sin embargo, al utilizar un mayor número de niveles de calidad, habría también que modificar el comportamiento del algoritmo en situaciones donde la tasa de bit disponible se reduce de manera abrupta, ya que en esos casos, es deseable reducir rápidamente la tasa de bit solicitada para adaptarse a las condiciones de la red. Si la conmutación a un nivel de calidad más bajo se produce de manera secuencial (una a una), al existir muchos niveles de calidad, la adaptación será lenta, dando lugar a mayores degradaciones en la calidad percibida. El conjunto de representaciones de vídeo (modeladas por su tasa de bit) que los usuarios pueden seleccionar es el siguiente: $Q = \{1; 1,5; 2; 4; 6; 8; 10; 12\} Mbps$

- Estrategia de adaptación basada en el concepto AIMD (Additive Increase Multiplicative Decrease):

  - Estimación de la tasa de bit en situaciones de congestión (reducción de nivel de calidad). Se ha cambiado la estrategia utilizada anteriormente, que consistía en disminuir el nivel de calidad en una unidad por cada nuevo segmento de vídeo solicitado mientras el buffer esté en la zona crítica. La nueva estrategia realiza una estimación de la tasa de bit disponible y reduce la tasa de bit solicitada en consecuencia. Esto permite que, tras detectar la congestión, el nivel de calidad se reduzca de manera más abrupta, no de unidad en unidad. El objetivo de esta modificación es intentar aliviar la congestión lo antes posible. Se debe destacar que mientras el buffer se encuentre en la zona crítica el nivel de calidad no se podrá aumentar (aunque la estimación de la tasa de bit lo permitiese). Además, la estimación de la tasa de bit se realiza utilizando únicamente información relativa al último fragmento de vídeo recibido. En general, el principal inconveniente que presenta esta medida es que al poder producirse cambios de calidad muy abruptos, la calidad percibida por los usuarios puede verse afectada, ya que el cambio de calidad será muy fácilmente detectable por los mismos.
  - Tras producirse un cambio a un nivel de calidad superior, no se podrá volver a aumentar el nivel de calidad solicitado hasta haber recibido un cierto número de fragmentos de vídeo, con el objetivo de que no se produzcan aumentos muy bruscos en el nivel de calidad solicitado.

Como se puede ver, en este experimento se está utilizando un algoritmo de adaptación relativamente conservador. En primer lugar se ha introducido un aumento considerable del tamaño del buffer, el cual se ha establecido en 30 segundos de vídeo (valor similar al que utiliza el reproductor Smooth Streaming). Esto hace que el margen de

tiempo con el que cuenta el algoritmo para adaptarse a las condiciones de la red aumente notablemente. En realidad, el tamaño del buffer de los players comerciales suele ser mayor (del orden de un par de minutos). Sin embargo, suelen planificar las peticiones de los fragmentos de vídeo para conseguir un nivel de estable de unos 30 segundos de vídeo. Por otro lado, se ha establecido una zona crítica muy conservadora (20 segundos), lo cual permite que el algoritmo reaccione rápidamente a degradaciones en el rendimiento de la red.

El aumento del número de representaciones del vídeo permite que el aumento de la demanda al conmutar a niveles de calidad superiores sea más escalonado. Como contrapartida, dicho aumento requiere introducir lógica adicional cuando la conmutación se realiza hacia niveles de calidad inferiores (motivada por situaciones de congestión en la red), con el objetivo de que la situación de congestión sea aliviada lo más rápido posible.

Los resultados obtenidos para el total de usuarios de vídeo agregados se muestran en la tabla 5.23.

Tabla 5.23: Resultados agregados del experimento de simulación 3

Simulación	**T buffering inicial (s)**		**T rebuffering (s)**		**N rebuffering**	
	Media	Desv. típica	Media	Desv. típica	Media	Desv. típica
3.1	4,728	3,520	0,000	0,000	0,000	0,000
3.2	4,194	3,207	0,000	0,000	0,000	0,000
3.3	4,230	3,162	0,000	0,000	0,000	0,000
3.4	2,326	1,829	0,000	0,000	0,000	0,000
3.5	2,061	1,692	0,000	0,000	0,000	0,000
3.6	2,037	1,685	0,000	0,000	0,000	0,000

Como se puede ver en los resultados agregados, las modificaciones que se han realizado al algoritmo han permitido que, mediante un aumento poco significativo en los tiempos de buffering inicial, no se produzcan eventos de rebuffering durante la reproducción del vídeo, con la mejora en la calidad percibida que ello supone.

A continuación se presentan los resultados del conjunto de experimentos realizado, en función de la capacidad de los enlaces de acceso de los usuarios (canales D_i). En estas tablas, en vez de presentar los resultados de los tiempos de rebuffering y eventos de rebuffering (no hay rebuffering, como se desprende de la tabla anterior de resultados agregados), se muestra el nivel de calidad que los clientes han solicitado. El nivel de calidad solicitado ha sido expresado mediante un identificador cuyo rango es $[0 \dots 7]$ y que se corresponde con las representaciones de calidad detalladas anteriormente.

En el primer experimento, correspondiente al escenario más limitado en cuanto a capacidad de red, se puede ver cómo todos los usuarios, excepto aquellos con conexiones de acceso de 2 Mbps, solicitan niveles de calidad similares, ya que la red no es capaz

Tabla 5.24: Resultados de la simulación 3.1

Canal D_i	**T buffering inicial (s)**		**Nivel de calidad solicitado**	
	Media	Desv. típica	Media	Desv. típica
2 Mbps	14,011	2,791	0,000	0,000
6 Mbps	5,789	2,561	1,510	0,502
20 Mbps	4,205	3,533	1,721	1,462
30 Mbps	3,631	3,688	1,855	1,435
100 Mbps	2,783	2,910	1,855	1,388

Tabla 5.25: Resultados de la simulación 3.2

Canal D_i	**T buffering inicial (s)**		**Nivel de calidad solicitado**	
	Media	Desv. típica	Media	Desv. típica
2 Mbps	13,646	1,035	0,000	0,000
6 Mbps	5,561	2,237	1,865	0,486
20 Mbps	3,472	3,006	1,969	1,513
30 Mbps	3,147	3,951	1,836	1,506
100 Mbps	2,593	3,334	2,010	1,493

Tabla 5.26: Resultados de la simulación 3.3

Canal D_i	**T buffering inicial (s)**		**Nivel de calidad solicitado**	
	Media	Desv. típica	Media	Desv. típica
2 Mbps	13,139	2,306	0,000	0,000
6 Mbps	5,691	2,396	1,866	0,485
20 Mbps	3,497	3,009	2,141	1,541
30 Mbps	2,800	2,646	2,148	1,553
100 Mbps	2,532	2,497	2,146	1,479

Tabla 5.27: Resultados de la simulación 3.4

Canal D_i	**T buffering inicial (s)**		**Nivel de calidad solicitado**	
	Media	Desv. típica	Media	Desv. típica
2 Mbps	11,459	0,311	0,000	0,000
6 Mbps	4,010	0,688	1,881	0,460
20 Mbps	1,425	0,790	3,428	2,220
30 Mbps	1,097	0,768	4,393	1,441
100 Mbps	0,713	0,772	4,445	2,295

Tabla 5.28: Resultados de la simulación 3.5

Canal D_i	**T buffering inicial (s)**		**Nivel de calidad solicitado**	
	Media	Desv. típica	Media	Desv. típica
2 Mbps	11,279	0,001	0,000	0,000
6 Mbps	3,792	0,152	1,882	0,459
20 Mbps	1,151	0,129	4,625	1,224
30 Mbps	0,753	0,001	6,393	1,795
100 Mbps	0,334	0,363	6,392	1,797

Tabla 5.29: Resultados de la simulación 3.6

Canal D_i	**T buffering inicial (s)**		**Nivel de calidad solicitado**	
	Media	Desv. típica	Media	Desv. típica
2 Mbps	11,279	0,000	0,000	0,000
6 Mbps	3,763	0,030	1,882	0,459
20 Mbps	1,134	0,055	4,626	1,224
30 Mbps	0,752	0,000	6,394	1,795
100 Mbps	0,229	0,000	6,394	1,795

de soportar tasas de bit mayores, por lo que la capacidad de los canales D_i no supone diferencia alguna. Para el caso de usuarios con 2 Mbps, en este caso sí existe limitación en el canal de acceso, por lo que no podrán solicitar calidades superiores al nivel 0 (nótese el factor reductor en la estimación de la tasa de bit disponible).

En los dos últimos experimentos mostrados (correspondientes a las simulaciones 3.5 y 3.6), se puede ver cómo cada grupo de usuarios solicita el nivel máximo que le permite su canal de acceso. En este análisis se deben tener presente las consideraciones que se plantearon anteriormente con respecto al algoritmo de adaptación de calidad y la estimación de la tasa de bit que dicho algoritmo aplica:

- La estimación de la tasa de bit se calcula utilizando la diferencia entre el instante en el que se solicita un fragmento de vídeo y el instante en el que se recibe dicho paquete, por lo que entran en juego los mecanismos que TCP impone para controlar la congestión, siendo la estimación algo más baja que el valor real.
- A la estimación de la tasa de bit ofrecida por la red se le aplica un factor reductor del 50 %.
- El algoritmo comienza solicitando el nivel de calidad mínimo, por lo que existe un periodo transitorio hasta alcanzar un nivel de calidad estable

Estas consideraciones condicionan los resultados obtenidos. Como ejemplo, en la figura 5.21 se muestra una traza correspondiente a un usuario con canal D_i=20Mbps, en la que se representa el nivel de calidad solicitado para los primeros 100 fragmentos de vídeo.

Como se puede ver en la figura, existe un cierto periodo transitorio en el que el nivel de calidad solicitado va aumentando hasta converger al nivel de calidad 5, correspondiente a una tasa de bit de 8 Mbps, que es la máxima a la que pueden optar los usuarios con canal D_i=20 Mbps con el algoritmo de adaptación utilizado. Podría pensarse que el nivel de calidad que deberían obtener los usuarios con canal D_i=20 Mbps es el nivel 6, correspondiente a 10 Mbps (la mitad de la capacidad de dicho canal). Sin embargo, debido al overhead de la torre de protocolos utilizada, la estimación de la tasa de bit que obtiene el algoritmo de adaptación (a nivel de aplicación) será menor que 10 Mbps.

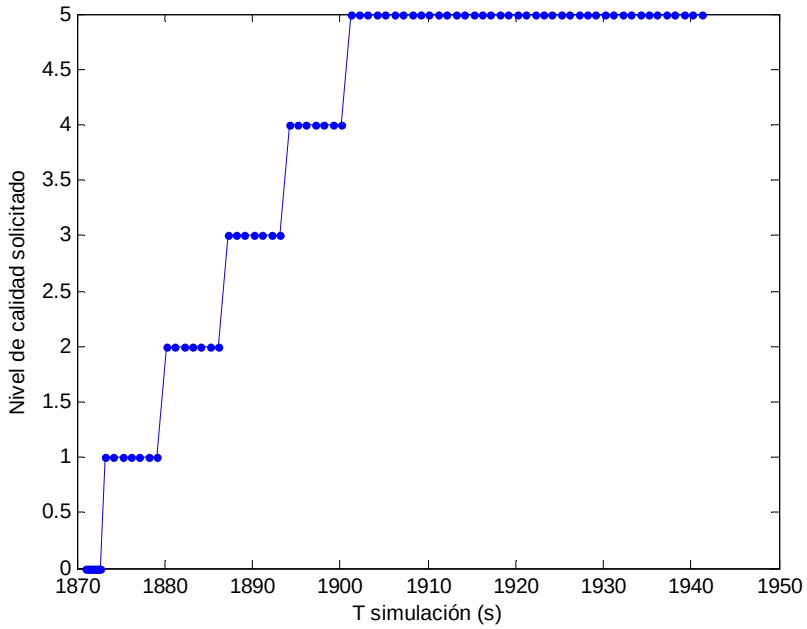


Figura 5.21: Traza del nivel de calidad solicitado por un usuario con canal D_i=20Mbps

5.4.2.4. Conclusiones extraídas de los experimentos de simulación

La implementación del algoritmo de adaptación de calidad juega un papel crucial en el rendimiento que se puede obtener de la red.

Un algoritmo de adaptación correctamente diseñado permite que, incluso en escenarios de red donde la tasa de bit disponible es muy baja, se puedan alcanzar valores de tiempos de rebuffering muy bajos o nulos, a cambio de incrementar el tiempo de buffering inicial.

Los algoritmos de adaptación no deben perder de vista los mecanismos de congestión de TCP, con el objetivo de evitar comportamientos indeseados. Por ejemplo, cuando una conexión TCP está inactiva un cierto tiempo y retoma el envío de paquetes, puede darse la situación de que el nuevo envío se realice en una fase de arranque lento (por vencimiento de timers), por lo que dicha conexión obtendrá una tasa de bit menor durante algún tiempo, afectando al rendimiento obtenido y, dependiendo de la implementación del algoritmo, a las estimaciones del ancho de banda ofrecido por la red. Estos periodos de inactividad son habituales cuando el buffer de recepción del cliente está lleno, ya que el cliente espera a que se consuma un fragmento para solicitar el siguiente. La combinación de este efecto, con los problemas derivados de la equidad de los algoritmos (descritos más adelante), pueden tener consecuencias inesperadas en situaciones de saturación de la red.

Considérese el siguiente ejemplo: dos usuarios compiten por el ancho de banda en un enlace común. El primer usuario U_1 tiene un canal de acceso de ancho de banda BW_1. El segundo usuario U_2 tiene un canal de acceso de BW_2, $BW_2 < BW_1$. El canal de agregación no tiene capacidad suficiente para soportar la máxima calidad de vídeo a la que pueden aspirar U_1 y U_2 (teniendo en cuenta la limitación de su canal de acceso). U_1 empieza a utilizar el servicio de vídeo, el canal agregado está a su entera disposición, por lo que pronto consigue alcanzar un nivel de calidad $Q_{1_{max}}$ estable y su buffer de recepción se llena. Dependiendo de la longitud de los fragmentos de vídeo, puede darse la situación en la que tras pedir cada fragmento, el servidor inicie la fase de arranque lento. Si U_2 empieza a utilizar el servicio, en el caso ideal cabría esperar que ambos compartiesen la capacidad del enlace agregado y se convergiera a un nivel de calidad similar. Así pues, U_2 empezará a solicitar fragmentos de vídeo, elevando la utilización del enlace compartido e irá aumentando su tasa de bit utilizada de acuerdo a los mecanismos de control de congestión de TCP. Por su parte, U_1 intentará solicitar un nuevo fragmento de vídeo, al nivel de calidad $Q_{1_{max}}$, cuyo tiempo de transferencia se verá afectado tanto por la demanda de U2 como por el arranque lento de TCP. Si U_1 realiza una estimación del ancho de banda utilizando el tiempo de transferencia de este fragmento, es probable que dicha estimación sea menor que las estimaciones que esté realizando U_2, ya que dicha estimación se verá afectada por el control de congestión de TCP. Este error en las estimaciones puede dar lugar a comportamientos inestables y a situaciones en las que U_1 solicite un nivel de calidad inferior a U_2, independientemente de que $BW_1 > BW_2$.

El tamaño del buffer tiene consecuencias directas en los picos de tráfico que la red tiene que soportar, ya que típicamente los clientes tratan de llenar el buffer lo más rápidamente posible, sobre todo al principio de la transmisión [Akhshabi et al., 2011]. Algunos players comerciales, una vez el buffer alcanza un cierto nivel de ocupación, abandonan esta estrategia avariciosa y planifican las peticiones de los fragmentos de vídeo para mantener el nivel del buffer estable [Akhshabi et al., 2012].

La equidad (fairness) del algoritmo es un aspecto muy importante. Como se analiza en [Akhshabi et al., 2011], players comerciales como "Smooth Streaming" no presentan comportamientos adecuados en situaciones en las que varios clientes compiten por el ancho de banda de un enlace saturado. Sin embargo, el carácter estocástico de este fenómeno, hace difícil su análisis, teniendo en cuenta que no se disponen de los detalles de implementación concretos del algoritmo Smooth Streaming.

Durante el desarrollo de esta sección de la tesis, en la que se ha ido depurando el algoritmo de adaptación, se obtuvieron algunos resultados que mostraban un cierto comportamiento no equitativo. Sin embargo, en las sucesivas iteraciones que se han realizado, dicho comportamiento se ha visto reducido. En cualquier caso, y como línea

de trabajo futuro, se podría añadir una componente de aleatoriedad que ayudase a incrementar todavía más la equidad del algoritmo, como se recomienda en [Gao et al., 2006] y en [Jiang et al., 2012].

Por último, se debe destacar que el tipo de tráfico que genera el streaming de vídeo adaptativo sobre TCP, en un escenario de hora cargada, hace que los enlaces de agregación (enlaces C_i en los experimentos llevados a cabo) tengan que estar correctamente dimensionados para soportar la demanda de todos los usuarios, ya que la ganancia estadística es limitada cuando la mayoría de los usuarios están utilizando el servicio.

5.5. Resumen y conclusiones

El streaming de vídeo basado en MPEG-DASH, y en general, el streaming de vídeo adaptativo transportado sobre protocolos fiables (como TCP), pueden introducir una serie de degradaciones en la calidad percibida del servicio. Más concretamente, los tiempos de espera e interrupciones, además de las variaciones en la calidad de vídeo (generalmente implementadas mediante variaciones en la tasa de bit de codificación del vídeo transmitido), afectan de manera negativa a la experiencia de usuario.

En este capítulo se ha llevado a cabo un estudio de esta degradación de calidad, planteando un modelo que estima el efecto sobre la calidad percibida del tiempo de buffering inicial, tiempo de rebuffering total, número de eventos de rebuffering y variaciones de calidad de vídeo.

Para desarrollar este modelo se han realizado un conjunto de experimentos de calidad subjetiva en los que una media de 20 voluntarios (tanto personas familiarizadas con la tecnología, como usuarios poco frecuentes de servicios de streaming de vídeo sobre Internet) evaluaron la calidad de diferentes vídeos, en los que se introdujeron de manera controlada distintas degradaciones. Estos experimentos fueron realizados mediante una plataforma web de evaluación de calidad de vídeo, adaptada a las necesidades de la tesis.

Tras analizar y procesar los datos recogidos en los experimentos, se han planteado modelos matemáticos que proporcionan una estimación de la degradación en la calidad percibida, recogidos en las ecuaciones 5.11, 5.12, 5.13 y en el algoritmo 1.

A la vista de estos modelos, se puede concluir que los eventos de rebuffering son la principal causa de degradación en la calidad percibida del streaming de vídeo adaptativo sobre TCP. Por otro lado, dado un tiempo de rebuffering total, afecta negativamente a la experiencia de usuario si dicho tiempo se reparte entre varios eventos de rebuffering. Así pues, es preferible, en términos de calidad, aumentar el tiempo de buffering inicial con el objetivo de reducir o eliminar los eventos de rebuffering. En cuanto al efecto de las variaciones en la calidad de vídeo, se ha propuesto un algoritmo capaz de explicar los resultados obtenidos en los experimentos subjetivos, teniendo en cuenta el valor medio

de calidad y el valor de calidad más alejado de dicha media. Cuando el valor de calidad más alejado de la media es mayor que ésta, se produce un incremento en la calidad percibida, mientras que si este valor es menor que la media, la calidad percibida se verá reducida. La diversidad en los resultados obtenidos pone de manifiesto la complejidad del efecto de las variaciones de calidad de vídeo, por lo que se plantea ampliar este estudio como parte del trabajo futuro.

Los efectos analizados en este capítulo son consecuencia directa de la incapacidad de la red para soportar la demanda de tráfico. Para analizar la relación entre la capacidad de la red, la demanda de tráfico y los tiempos de buffering inicial y de rebuffering, se han llevado a cabo una serie de simulaciones de red. Estas simulaciones han corroborado el compromiso entre tiempo de buffering inicial y tiempo de rebuffering, comentando anteriormente, además de poner de manifiesto la importancia del diseño del algoritmo de adaptación de calidad de vídeo utilizado.

Capítulo 6

Conclusiones y líneas de trabajo futuras

En este capítulo se extraen las principales conclusiones de esta tesis, mediante el análisis de los objetivos que se marcaron en la sección 1.2. Además se analiza el marco de trabajo en el que se ha realizado esta tesis doctoral. Por último, se plantean una serie de líneas futuras de investigación con las que continuar la labor comenzada en este trabajo.

6.1. Análisis de los objetivos

6.1.1. Propuesta de un modelo global de estimación calidad percibida para servicios de streaming de vídeo adaptativo OTT

En el capítulo 3 se propone un modelo para la estimación de la calidad percibida global en servicios de streaming de vídeo OTT. Partiendo de la descripción de dicho servicio (realizada mediante el modelo de descripción de servicios propuesto), se plantea un modelo que combina las aportaciones a la calidad de cada uno de los principales componentes del servicio. Más concretamente, el modelo tiene en cuenta los siguientes aspectos del servicio:

- Calidad de vídeo
- Calidad de audio
- Calidad (o degradación) asociada a la sincronización entre el audio y el vídeo (en inglés, lip-sync)
- Degradación asociada al efecto de la red (degradación por transmisión).
- Tiempo de seeking o acceso aleatorio

- Tiempo de cambio de canal

Para agregar las contribuciones a la calidad total de cada componente del modelo, en esta tesis se distingue entre los componentes que se han denominado "componentes continuos" y "componentes puntuales". Los componentes continuos son aquellos cuyo efecto está presente durante la mayor parte del tiempo de prestación del servicio (vídeo, audio, etc.), mientras que los componentes puntuales son aquellos cuyo efecto solo aplica en intervalos de tiempo limitados (cambio de canal, etc.). En el modelo propuesto, la influencia de los componentes puntuales depende de la calidad de los componentes continuos, de acuerdo a las siguientes reglas:

- La influencia de la calidad asociada a las componentes puntuales es relevante para el cómputo de la calidad total solo si la calidad de la totalidad de los componentes continuos alcanza un cierto umbral. Es decir, si la calidad de los componentes continuos es baja, el nivel de calidad de los componentes puntuales es poco relevante.

- La influencia de la calidad asociada a las componentes puntuales puede ser moderada si la calidad de la totalidad de los componentes supera un cierto valor. Es decir, si la calidad de los componentes continuos es muy alta, la tolerancia en cuanto a la calidad de los componentes puntuales (que afectan durante una fracción de tiempo pequeña) puede ser mayor, es decir, su relevancia puede verse moderada.

Así pues, de manera genérica, el modelo global de calidad tiene la siguiente forma:

$$Q = \sum_{i=1}^{N_c} c_i \cdot Q_{c_i} + \sum_{j=1}^{N_p} p_j \cdot Q_{p_j} = Q_C + \sum_{j=1}^{N_p} p_j \cdot Q_{p_j} \tag{6.1}$$

El factor Q_C, asociado a los componentes continuos se ha definido mediante la siguiente ecuación:

$$Q_C = Q_{av_{total}} - I_{tra} = Q_{av} - I_{ls} - I_{tra} \tag{6.2}$$

Como se puede ver, se ha introducido un factor de calidad Q_{av}, el cual cuantifica la calidad audiovisual, suponiendo sincronización perfecta entre los flujos de audio y vídeo y ausencia de degradaciones asociadas a la red y a los mecanismos de transmisión. El efecto de la falta de sincronización entre los flujos se recoge en el factor de degradación I_{ls}, mientras que el efecto de la red se modela mediante I_{tra}.

La expresión propuesta para modelar Q_{av} está basada en [Garcia et al., 2013] y en la recomendación ITU-T P.1201.2 [ITU, 2012e], y se calcula a partir de la calidad del

audio y del vídeo, utilizando los parámetros de ajuste que se proporcionaron en la tabla 3.9.

$$Q_{av} = 0,7 \cdot (\alpha + \gamma \cdot Q_v + \mu \cdot Q_a \cdot Q_v) + 0,3 \cdot (a - b \cdot Icod_a - c \cdot Icod_v) \tag{6.3}$$

En cuanto a la calidad de vídeo, la recomendación ITU-T P.1201.2 propone un modelo de estimación que utiliza únicamente información contenida en las cabeceras de los paquetes de los flujos de transporte de vídeo. Aunque este es un enfoque que permite llevar a cabo estimaciones de calidad de manera eficiente (en cuanto a tiempo de cómputo), al no analizar el contenido de las tramas decodificadas se está "desperdiciando" información valiosa para realizar la estimación de la calidad percibida. Esto ha motivado que en esta tesis se desarrolle un nuevo modelo de estimación de calidad de vídeo sin referencia, el cual se lleva a cabo en el capítulo 4 y se resume en la sección 6.1.2.

En cuanto a la degradación en el audio, el modelo que se utilizará en esta tesis es el recomendado por ITU-T P.1201.2:

$$Icod_a = a_{1a} \cdot e^{a_{2a} \cdot BitRate} + a_{3a} \tag{6.4}$$

Para estimar la degradación en la calidad percibida que supone la falta de sincronización entre los flujos de audio y vídeo se ha propuesto el siguiente modelo:

$$I_{ls} = \begin{cases} 100, & T \leq A_1 \\ \alpha \cdot log(-T) + \beta, & A_1 < T < D_1 \\ 0, & D_1 \leq T \leq D_2 \\ \gamma \cdot log(T) + \xi, & D_2 < T < A_2 \\ 100, & T \geq A_2 \end{cases} \tag{6.5}$$

Este modelo se ha expresado de forma paramétrica, con el objetivo de tener en cuenta la dependencia de los umbrales de aceptabilidad y detección con respecto al tipo de contenido de la secuencia de vídeo. Los umbrales propuestos se pueden consultar en la tabla 3.13.

Los efectos que pueden introducir en la calidad percibida las condiciones de la red y los protocolos utilizados para transportar el vídeo se analizan en el capítulo 5 y se resumen en la sección 6.1.3.

En cuanto a los componentes puntuales contemplados en el modelo: cambio de canal y acceso aleatorio, los modelos propuestos son los siguientes:

Para cuantificar el efecto del tiempo del cambio de canal en la calidad percibida se propone la utilización del modelo presentado en [Kooij et al., 2009b]. Dicho modelo propone una expresión para estimar la MOS en función del tiempo de cambio de canal:

$$MOS_{z,var=0} = \begin{cases} -2,1 \cdot T_z + 4,9, & 0 \leq T_z \leq 1,04 \\ -1,067 \cdot ln(T_z) + 2,757, & 1,04 \leq T_z \leq 4,97 \\ 1,05, & 4,97 \leq T_z \end{cases} \quad (6.6)$$

De manera análoga, para el caso de la estimación de la calidad asociada al acceso aleatorio, se propone una expresión similar, a la que se le ha añadido un factor adicional para incluir una penalización en la calidad percibida en caso de que el acceso aleatorio no se realice de manera precisa.

6.1.2. Propuesta de un modelo de estimación de calidad percibida de vídeo

En esta tesis se ha llevado a cabo el desarrollo de un modelo sin referencia para la estimación de calidad percibida en vídeo. Dicho modelo está orientado a contenidos con resolución Full-HD (1920x1080) codificados en H.264/AVC. Como se ha comentado anteriormente, las degradaciones que introduce la red serán analizadas de manera independiente, por lo que el modelo de estimación de calidad de vídeo se centra en los defectos que se hayan podido introducir en el proceso de codificación.

Más concretamente, el modelo propuesto tiene como objetivo obtener una estimación de VQM_VFD sin utilizar la señal de vídeo original, basándose únicamente en características del vídeo recibido. VQM_VFD es el resultado que genera el modelo de referencia completa propuesto en [Wolf and Pinson, 2011], el cual es una evolución del modelo de VQM de NTIA estandarizado en ITU-T J.144 [ITU, 2004c], adaptado a resoluciones más altas y a nuevos tipos de degradaciones.

Para desarrollar el modelo, en primer lugar se seleccionaron un conjunto de secuencias de vídeo, se codificaron a distintas tasas de bit y se calculó el valor de VQM_VFD para cada una de ellas (utilizando la secuencia de vídeo original y la secuencia de vídeo degradada, es decir, codificada/comprimida en H.264). Estos valores de VQM_VFD en función de la tasa de bit de codificación constituyen el conjunto de datos de entrenamiento del modelo. Tras analizar dichos datos, se puso de manifiesto que se puede expresar la dependencia de VQM_VFD con respecto a la tasa de bit de codificación mediante la siguiente ecuación:

$$VQM\_VFD = a \cdot bitRate^b \quad (6.7)$$

La siguiente fase del desarrollo consistió en decidir cómo ajustar un modelo mate-

mático capaz de estimar los parámetros a y b para cada secuencia de vídeo del conjunto de entrenamiento. Tras analizar varias posibilidades, se optó por desarrollar el modelo mediante la utilización de una red neuronal, la cual utiliza como parámetros de entrada las siguientes variables:

- Información espacial, SI
- Información temporal, TI
- Información espacial media, ASI
- Información temporal media, ATI
- Entropía media, H_{avg}
- Entropía máxima, H_{max}
- Información temporal media de bordes, ATI-Sobel
- Variación sobre la información temporal media de bordes, ATI-Sobel-2
- Módulo medio de los vectores de movimiento, μ_{MVM}
- Coherencia del movimiento, σ_{DVM}
- Cociente entre el módulo medio y la coherencia del movimiento μ_{MVM}/σ_{DVM}

En general, todas estas variables representan diferentes características de la complejidad espacial y temporal de la secuencia. Para más información al respecto se puede consultar la sección 4.3.4.3.

El entrenamiento de la red neuronal se llevó a cabo utilizando dos técnicas diferentes (algoritmo Levenberg-Marquardt y regularización bayesiana), obteniendo resultados similares en ambos casos. El MSE obtenido en la predicción de los parámetros a y b fue de 0,0074 y 0,0041 para el algoritmo Levenberg-Marquardt y regularización bayesiana respectivamente. En la sección 4.3.5 se muestran más detalles en cuanto a los resultados obtenidos.

6.1.3. Propuesta de un modelo de estimación de degradación en la calidad percibida asociada a la red y a los mecanismos de transmisión

En el capítulo 5 se ha llevado a cabo el análisis de las degradaciones que se pueden producir en la calidad percibida como consecuencia de transmitir el flujo audiovisual a través de la red, utilizando mecanismos de streaming adaptativo sobre HTTP. En concreto, se han estudiado los siguientes aspectos:

- Tiempo de buffering inicial
- Número de eventos de rebuffering
- Tiempo total de los eventos de rebuffering
- Cambios en la calidad de vídeo motivados por los algoritmos de adaptación.

El análisis de cada uno de estos puntos se ha llevado a cabo siguiendo una metodología similar, la cual se basa en la obtención de datos de valoraciones de calidad percibida de usuarios reales. La obtención de dichos datos se ha llevado a cabo mediante la utilización de una plataforma web de evaluación de calidad de vídeo, QualityCrowd2, la cual ha sido adaptada las necesidades de la tesis, tal y como se describe en el apéndice C. Estas evaluaciones subjetivas sirven para entender la dependencia que tiene la calidad percibida con respecto a las variables de estudio.

Los modelos propuestos son los siguientes:

Degradación de la calidad percibida asociada al tiempo de buffering inicial:

$$I_{Tbuffering\,inicial} = a \cdot \sqrt{b \cdot T_{buffering\,inicial}} + c \tag{6.8}$$

Degradación de la calidad percibida asociada al tiempo de rebuffering:

$$I_{Trebuffering} = \frac{a \cdot T_{rebuffering}}{1 + b \cdot T_{rebuffering}} \tag{6.9}$$

Degradación de la calidad percibida asociada al número de eventos rebuffering:

$$I_{Nrebuffering} = a \cdot (1 - N^{b}_{rebuffering}) \tag{6.10}$$

En cuanto al efecto de los cambios en la calidad del vídeo, se ha propuesto un algoritmo que modela los hallazgos descubiertos en las pruebas de evaluación de calidad subjetiva realizadas. Estas pruebas han puesto de manifiesto que los cambios en la calidad de vídeo pueden afectar tanto negativa como positivamente en la valoración global de la calidad percibida, en función de los niveles de calidad más "atípicos" que se produzcan a lo largo de la reproducción. Por ejemplo, si a lo largo de la reproducción del vídeo predominan niveles de calidad altos, cuando se conmuta a un nivel de calidad bajo, se produce una degradación adicional en la calidad percibida. El caso contrario también aplica: si los niveles de calidad suelen ser bajos, el que se conmute a un nivel de calidad superior se recompensa en las valoraciones de calidad de los usuarios.

Por último, se han llevado a cabo una serie de simulaciones de red con el objetivo de analizar la dependencia de las variables anteriores con respecto a la capacidad de la red y de los algoritmos de adaptación utilizados.

6.2. Difusión de resultados

Las ideas y contribuciones de esta tesis han sido de gran utilidad en el desarrollo del proyecto de investigación "VideoXperience: Mejora Efectiva de la Experiencia de Usuario en la Nueva Era de Servicios Digitales mediante la Provisión de nuevas Tecnologías de Supercompresión en Streaming". Este proyecto forma parte del subprograma INNPACTO del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2008-2011, y está financiado por el Ministerio de Ciencia e Innovación, actual Ministerio de Economía y Competitividad. Los objetivos primordiales del proyecto son dos:

- Caracterizar el dimensionamiento de Internet para poder ofrecer servicios de vídeo de alta calidad con una experiencia de usuario medible y similar a los actuales sistemas de TDT e IPTV desplegados por operadores.

- Cubrir el gap existente entre los resultados obtenidos con el primer objetivo y la capacidad de las redes actuales. Para ello se desarrollará un nuevo sistema de codificación de imagen y video capaz de satisfacer dicha experiencia de usuario en Internet sobre cualquier red de acceso fija o móvil. Esto reducirá el coste por byte, aumentará la capacidad de las redes existentes y mejorará la experiencia de usuario.

En el contexto de este proyecto se ha realizado difusión de resultados mediante las siguientes publicaciones:

Pedro de la Cruz, Joaquín Navarro, Raquel Pérez, Francisco González. **Estimating Perceived Video Quality from Objective Parameters in Video over IP Services**. En 7th IARIA International Conference on Digital Telecommunications, ICDT 2012, pp. 65–68.

Jose Javier García Aranda, Marina González Casquete, Mario Cao Cueto, Joaquín Navarro Salmerón, Francisco González Vidal. **Logarithmical hopping encoding: a low computational complexity algorithm for image compression**. Aceptado para publicación en IET Image Processing Journal.

6.3. Líneas de trabajo futuro

A lo largo del desarrollo de esta tesis doctoral se han identificado varias líneas de trabajo con las que continuar, complementar y aplicar las contribuciones de la misma:

- Validación de la función propuesta como factor de peso para los componentes puntuales $f(Q_C)$ (ecuación 3.13).

- Validación de la función propuesta para cuantificar la degradación en la calidad asociada al error en el acceso aleatorio en vídeo (ecuación 3.47).
- Ampliación de la simulaciones llevadas a cabo en la sección 5.4.2, incluyendo nuevos algoritmos de adaptación de calidad de vídeo.
- Diseño, desarrollo y prueba de una arquitectura de monitorización (y control) de calidad percibida en servicios de streaming de vídeo OTT: esta línea de trabajo tiene como principal objetivo la aplicación de los modelos propuestos en esta tesis para el desarrollo de una solución de monitorización de calidad percibida por los usuarios. Así pues, sería necesario llevar a cabo la implementación de los modelos propuestos en diferentes dispositivos de cliente (set-top boxes, dispositivos móviles, librerías Javascript para clientes web, etc.) y el diseño de la arquitectura de recogida y análisis de los datos generados en los clientes, contemplando técnicas de minería de datos, visual analytics, etc. Esta arquitectura de monitorización podría constituir un servicio independiente de los proveedores de contenido, ofreciendo librerías que éstos tendrían que integrar en sus clientes para tener acceso a las estimaciones de QoE.
- Ampliación del estudio del efecto de los algoritmos de adaptación de calidad en la calidad percibida: en los resultados obtenidos en esta tesis se pone de manifiesto la complejidad asociada a la estimación del efecto que los cambios de calidad de vídeo tienen sobre la calidad percibida. Este problema tiene entidad suficiente para constituir una línea de investigación independiente, que analice en profundidad las distintas dimensiones del problema (número de cambios de calidad, diferencia entre niveles de calidad, efecto memoria, etc.).
- Adaptación del modelo de calidad de vídeo al nuevo estándar H.265 [ITU, 2013] y a nuevas resoluciones (4K UHD).
- Estudio de nuevos enfoques para la estimación de calidad de vídeo, basados por ejemplo en técnicas de reconocimiento automático de imágenes, extracción de información sobre el contenido, etc.

Apéndice A

Modelo de descripción de servicios

A.1. Introducción y motivación

En esta sección se introduce un modelo de descripción de servicios basado en componentes de servicio y funciones reutilizables. Este modelo tiene los siguientes objetivos:

- Plantear un marco común y formal en el análisis de servicios de vídeo.
- Servir como herramienta a la hora del diseño de modelos de estimación de calidad percibida.

El primer objetivo está motivado por un aspecto que se ha puesto de manifiesto al realizar el estudio del estado del arte y es que en la literatura actual es común la utilización de un amplio abanico de términos para hacer referencia a servicios de vídeo, confundiendo en muchos casos el servicio en sí mismo (desde el punto de vista del usuario) con la implementación o la plataforma tecnológica que se utiliza para desplegar dicho servicio. Uno de los casos más típicos es la utilización del término IPTV como sinónimo del servicio de difusión de televisión o televisión lineal, cuando realmente IPTV es un sistema o una implementación concreta del servicio de difusión de televisión. El modelo de descripción de servicios que se propone intenta poner de manifiesto el hecho de que, para un usuario final, el servicio de difusión de televisión debería ser indistinguible tanto si éste es ofrecido mediante una red Internet Protocol (IP) gestionada por un operador, como si es ofrecido mediante una plataforma de vídeo OTT, como si es ofrecido mediante la difusión de ondas electromagnéticas por el aire.

En cuanto al segundo objetivo (de especial interés para esta tesis), al representar un servicio mediante el modelo que se propone en este capítulo, se pondrán de manifiesto sus distintos componentes, cada uno de los cuales contribuirá en cierta medida a la

calidad percibida por el usuario. Esta representación basada en componentes será de gran utilidad a la hora de plantear modelos de estimación de calidad percibida en servicios de vídeo, ya que la calidad percibida "total" podrá ser expresada en función de la calidad de cada uno de los componentes del servicio.

A.2. Marco de referencia

En esta sección se enumeran los estándares y recomendaciones en los que se apoya el modelo definido, destacando aquellos elementos o características que se han aplicado en el modelo.

ITU-T I.130 Método de caracterización de los servicios de telecomunicación soportados por una Red Digital de Servicios Integrados (RDSI) y de las capacidades de red de una RDSI [ITU, 1989].

Los objetivos principales de esta recomendación son:

- Proporcionar un entorno común y las herramientas necesarias para describir servicios.
- Mostrar cómo partiendo de la definición formal de un servicio se pueden definir protocolos y recursos de red para proveer dichos servicios.
- Hacer referencia a aquellas recomendaciones pertinentes a los dos puntos anteriores.

Esta recomendación define un método para caracterizar servicios que se divide en tres fases de actividad:

1. Descripción del servicio desde el punto de vista del usuario.
2. Descripción de la organización de funciones de red, en las que se asocian requisitos de servicio con capacidades de red.
3. Definición de capacidades de conmutación y señalización requeridas para dar soporte a los servicios definidos en la primera fase.

ITU-T I.140 Técnica basada en atributos para la caracterización de los servicios de telecomunicación soportados por una RDSI y de las capacidades de red de una RDSI [ITU, 1993c].

Esta recomendación describe una técnica para describir atributos y listas de valores de atributo. Además, este documento contiene una librería de atributos y valores de atributo utilizados en otras recomendaciones de la serie I de la ITU.

En esta recomendación se contemplan varios tipos de atributos:

- Dominantes: definen un subconjunto que contiene objetos similares y a este subconjunto se le denomina clase o categoría.
- Secundarios: definen un objeto particular.
- De cualificación: definen variantes de un objeto.

Además cada atributo ha de cumplir una serie de reglas descritas en el documento, como son por ejemplo:

- Tener un nombre y descripción asignados.
- Los atributos y sus valores pueden ser usados más de una vez por diferentes servicios o componentes de servicio.
- Cada atributo ha de ser descrito en base a tres perspectivas: genérica, servicio y red. La recomendación incluye una extensa lista con atributos y posibles valores de cada una de las tres perspectivas.

UML Lenguaje unificado de modelado.

Unified Modeling Language (UML) es el lenguaje de modelado de propósito general más conocido y más utilizado en el campo de la ingeniería del software orientado a objetos. Este estándar fue creado y se encuentra gestionado por el Object Management Group (OMG). UML incluye un conjunto de técnicas de notación gráfica para crear modelos visuales de sistemas de software orientados a objetos [Fowler and Scott, 1997].

Open IPTV Forum Service and Platform Requirements.

Este documento [Open IPTV Forum, 2008a] define unos requisitos de servicio y plataforma para una solución IPTV, ya sea basada en un modelo de red gestionado o abierto. Este documento presenta una lista muy completa y genérica de requisitos, ya que distingue entre requisitos de carácter obligatorio, recomendado u opcional.

Open IPTV Forum Open IPTV Forum Services and Functions for Release 2.

Este documento [Open IPTV Forum, 2008b] describe aquellos servicios y funcionalidades que han de estar presentes en las soluciones que sigan las especificaciones del IPTV Forum. En primer lugar se describen servicios genéricos, como pueden ser los servicios de televisión tradicionales, hasta servicios más complejos y novedosos, como son los servicios de comunicación integrados en una solución IPTV (chat, presencia, videoconferencias, etc.).

A.3. Descripción del modelo

En esta sección se describe de manera detallada el modelo propuesto de descripción de servicios basado componentes de servicio y funciones reutilizables. En primer lugar se enumeran los requisitos que debe cumplir dicho modelo, después se detallan los elementos que lo componen y por último se muestra una descripción gráfica del mismo mediante UML.

A.3.1. Objetivos

El propósito principal de este modelo es describir formalmente un servicio a partir de la percepción del mismo desde el punto de vista del usuario mediante la identificación de sus componentes en diferentes niveles de abstracción.

Aunque los objetivos generales del modelo fueron descritos en la introducción de este capítulo, a continuación se enumeran un conjunto de objetivos más detallados:

- Proporcionar un modelo detallado del servicio partiendo desde el punto de vista del usuario.
- Proporcionar diferentes niveles de abstracción en la descripción de un servicio, que vayan desde la percepción del usuario hasta los detalles de implementación.
- Proporcionar las herramientas de modelado necesarias para generar diagramas (representaciones gráficas) de descripción de servicio.
- Permitir la reutilización de elementos del modelo para la composición de servicios más complejos.

A.3.2. Elementos del modelo

El modelo de descripción de servicios está formado por diferentes elementos, los cuales se definen a continuación de mayor a menor nivel de abstracción, o dicho de otra forma, desde un punto de vista más cercano al usuario a un punto de vista más cercano a la implementación.

Servicio El elemento de mayor nivel de abstracción del modelo es el servicio. Un servicio se puede definir como un conjunto de actividades que buscan responder a las necesidades de un usuario y mediante el cual se hace una "entrega" de un producto intangible. Esta definición forma parte de la definición más amplia de servicio dada en la recomendación ISO 9000 [ISO, 2005a], que establece que un servicio es el resultado de llevar a cabo necesariamente al menos una actividad en la interfaz entre el proveedor y el cliente, la cual generalmente es intangible. Por ejemplo, mediante el servicio de

vídeo bajo demanda se lleva cabo una actividad entre proveedor y cliente en forma de alquiler de contenidos multimedia. En este modelo, el servicio es el elemento de mayor nivel, el cual será descrito un función de los elementos que se definen a continuación.

Sub-servicio Un sub-servicio es un elemento del modelo que se define como un conjunto de actividades que pertenecen a otro servicio de índole más general y que no pueden ser divididas en servicios más elementales, ya que dicha división no sería vista por el usuario como una actividad con el suficiente valor individual. La utilidad, y a veces la existencia, de estos sub-servicios puede estar condicionada a la de otros sub-servicios, en casos en los que se requieran éstos para poder satisfacer las necesidades del usuario.

Ejemplo: la guía de contenidos o de programación de un servicio de televisión lineal o de VoD se puede ver como un sub-servicio. En este caso, el sub-servicio de guía de contenidos/programación, encargado de mostrar la parrilla de contenidos/programación a un usuario, no puede ser dividido en sub-servicios más elementales y que tengan entidad en sí mismos (desde el punto de vista del usuario).

Componente de servicio Los componentes de servicio son bloques que representan operaciones básicas de un sub-servicio. Cada componente de servicio queda definido por un conjunto de parámetros y un conjunto de métodos u operaciones. Los parámetros se utilizan para concretar la configuración del componente de servicio y los métodos se utilizan para definir las operaciones concretas que el componente de servicio es capaz de realizar.

Ejemplo: siguiendo con el caso del sub-servicio guía de contenidos/programación, un posible componente de servicio es la "Visualización de listas de contenido". Este componente tendrá como parámetros los siguientes elementos (entre otros): número de elementos por página, tipo de presentación (mediante filas, mediante iconos, etc.), tiempo de respuesta en la paginación, etc. En cuanto a los métodos de este componente de servicio, los más representativos serían: métodos de navegación (ir a página siguiente, ir a página anterior, etc.), seleccionar elemento, aplicar filtros, etc.

Por el carácter genérico de los componentes de servicio, puede darse el caso de que varios sub-servicios compartan componentes de servicio. El componente de servicio es el elemento que aporta el carácter reutilizable al modelo. Así pues, con un conjunto suficientemente grande de componentes de servicio genéricos, se podrán definir múltiples servicios en base a ellos.

Ejemplo: el componente de servicio "control remoto" es un componente que puede formar parte de diferentes sub-servicios (como pueden ser la guía de contenidos/programación, el teletexto, la reproducción de contenido, entre otros), permitiendo al usuario realizar acciones de manera remota mediante algún tipo de terminal inalámbrico.

Bloques arquitecturales Dentro del modelo de descripción de servicios, los bloques arquitecturales representan un nivel de abstracción intermedio entre los componentes de servicio y la implementación concreta de dichos componentes. Así pues, los bloques arquitecturales se utilizan para concretar la estructura (arquitectura) de la implementación de cada componente de servicio. Expresado de otra manera, aprovechando términos propios de la ingeniería del software orientado a objetos, se pueden comparar los bloques arquitecturales con la definición de "interfaces" que son implementadas por el siguiente nivel del modelo.

En general, los bloques arquitecturales pueden agruparse en o pueden pertener a los siguientes conjuntos:

- Equipamiento de usuario.
- Red.
- Contenido.
- Gestión del servicio.

Ejemplo: el bloque arquitectural "interfaz inalámbrica" (de tipo "Equipamiento de usuario") es necesario para construir el componente de servicio "control remoto". Nótese que en la definición de bloque arquitectural no se incluyen los detalles de implementación, los cuales son tenidos en cuenta por el siguiente nivel del modelo.

Como se desprende de los ejemplos, el bloque arquitectural detalla la estructura de un componente de servicio, pero sin entrar en el detalle concreto de la implementación del mismo.

Implementación La implementación designa el medio o la forma mediante el cual se concretan o se desarrollan los bloques arquitecturales. Para que un sistema o solución pueda prestar un servicio descrito con este modelo, debe contar con implementaciones de todos los bloques arquitecturales que lo componen. Es importante destacar que para un mismo sistema o solución, pueden existen múltiples implementaciones. Cabe destacar también que, como la implementación representa la realización concreta de un bloque arquitectural, las implementaciones se pueden clasificar en los mismos grupos que los bloques arquitecturales (equipamiento de usuario, red, contenido y gestión del servicio).

Ejemplo: para satisfacer el bloque arquitectural "interfaz inalámbrica" para el componente de servicio "control remoto" se requiere una implementación de equipamiento de usuario, como por ejemplo, un emisor y un receptor de infrarrojos o un emisor y un receptor bluetooth.

En el apartado A.4 se proporciona una metodología para la aplicación de este modelo al dominio de los servicios multimedia, la cual contribuirá a la comprensión de las ideas que se han expuesto en este apartado.

A.3.3. Representación gráfica

Para la representación gráfica de servicios, se ha optado por la herramienta de modelado UML. Más concretamente, los servicios podrán ser descritos mediante diagramas de clase, aprovechando los mecanismos de herencia, asociación, composición, implementación de interfaces, etc. La figura A.1 muestra el diagrama de clases de un servicio genérico, donde se pueden apreciar las relaciones entre los distintos elementos de un servicio.

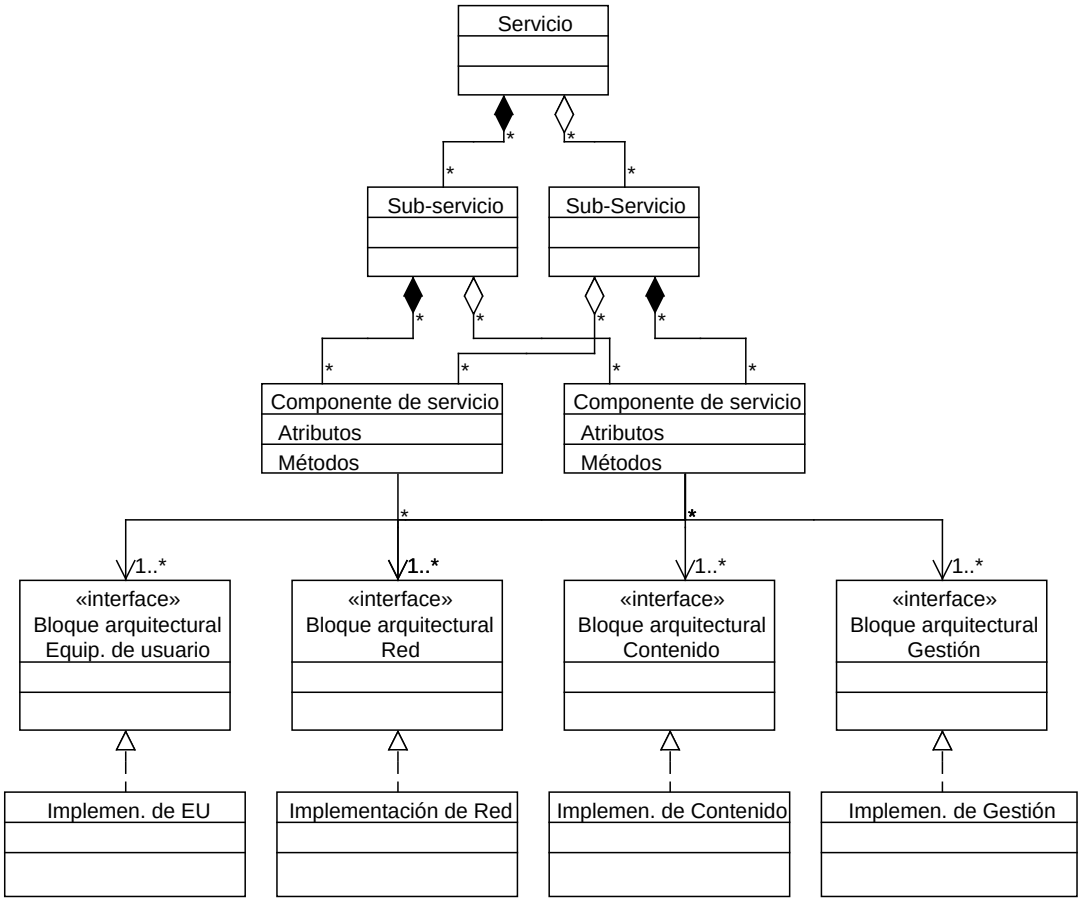


Figura A.1: Diagrama UML del modelo de descripción de servicios

Como se puede en ver la figura A.1, cada elemento del modelo representa un nivel de abstracción diferente. El primero de estos niveles corresponde al servicio. Un servicio

está compuesto por una serie de sub-servicios. Algunos de estos sub-servicios serán obligatorios y otros serán opcionales, como se refleja en el tipo de conector usado entre la clase Servicio y Sub-servicio. Las multiplicidades representan el carácter reutilizable de los sub-servicios, permitiendo que un sub-servicio forme parte de múltiples servicios.

Cada sub-servicio se compone de una serie de componentes de servicio, los cuales pueden ser obligatorios u opcionales. Como se puede ver, cada uno de estos componentes de servicio cuenta con un conjunto de parámetros y métodos que lo definen. De nuevo las multiplicidades ponen de manifiesto el carácter reutilizable de los componentes de servicio.

En el siguiente nivel se encuentran los bloques arquitecturales, los cuales se pueden ver como interfaces desde el punto de vista de UML, ya que establecen las funcionalidades de cada componente de servicio, pero sin especificar su implementación. Por último, cada una de estas interfaces (bloques arquitecturales) cuenta con una implementación concreta.

Como se puede ver, el modelo intenta ofrecer una visión completa del servicio, partiendo desde el punto de vista del usuario, incluyendo mayor detalle en cada nivel, hasta llegar a la implementación.

A.4. Metodología para la aplicación del modelo de descripción de servicios al dominio de los servicios multimedia

Una vez introducido el modelo de descripción de servicios, en este apartado se describe una metodología que permite la aplicación de dicho modelo al caso concreto de los servicios multimedia. El resto de la sección se organiza de la siguiente manera: en primer lugar, se introducen los pasos de los que consta la metodología y tras ello, se aplica dicha metodología al caso de dos de los servicios más representativos del dominio de los servicios multimedia, como son la televisión lineal y el vídeo bajo demanda VoD.

A.4.1. Descripción de la metodología

La metodología que se propone para la aplicación del modelo de descripción de servicios está basada en un enfoque top-down (de arriba abajo), alineado con la propia concepción del modelo, que permita, partiendo desde un punto de vista cercano a la percepción del usuario (alto nivel de abstracción), ir incluyendo en cada nivel del modelo más detalles hasta acercarse al nivel de la implementación (bajo nivel de abstracción).

Por tanto, los pasos de los que consta esta metodología son los siguientes:

1. Descomposición del servicio en sub-servicios.

2. Identificación de los componentes de servicio que se requieren para definir a los sub-servicios.

3. Descripción de la arquitectura de los componentes de servicio mediante la definición de bloques arquitecturales.

4. Concretar la implementación de cada uno de los bloques arquitecturales definidos en el paso anterior.

Con el objetivo de ilustrar esta metodología, a continuación se aplica al caso concreto de los servicios de televisión lineal y vídeo bajo demanda.

A.4.1.1. Aplicación de la metodología al servicio de televisión lineal

El servicio de televisión lineal es un servicio de audio y vídeo, muy popular y ampliamente extendido, donde el contenido que se puede consumir está prefijado en forma de diferentes canales de televisión (flujos multimedia) los cuales generalmente son recibidos por todos los usuarios del sistema de manera simultánea (broadcast).

En la figura A.2 se muestra la representación gráfica del servicio de televisión lineal, la cual se analiza en las siguientes líneas.

Paso 1: descomposición en sub-servicios Tras analizar las particularidades del servicio de televisión lineal, dicho sistema se puede ver como la combinación de los siguiente sub-servicios:

- Reproducción de contenido: un elemento básico, factor común de todos los servicios de vídeo. Este sub-servicio es una pieza indispensable en todos los servicios de vídeo ya que su misión es reproducir contenidos de audio y vídeo.

- Guía de canales o contenido (opcional): la guía de contenido es un sub-servicio presente (aunque con variaciones) en la mayoría de servicios de vídeo actuales. Presenta de forma ordenada el contenido que el usuario tiene disponible, permitiéndole obtener información detallada o acceder, si así lo desea, a dicho contenido.

- Teletexto: el teletexto es un sub-servicio textual que se emite junto al contenido del servicio de vídeo (generalmente servicios de televisión convencional). El teletexto ofrece un conjunto de páginas ordenadas numéricamente las cuales proporcionan información de diversa índole.

Paso 2: identificación de componentes de servicio Cada uno de los sub-servicios identificados en el paso anterior se puede descomponer a su vez en un conjunto de componentes de servicio.

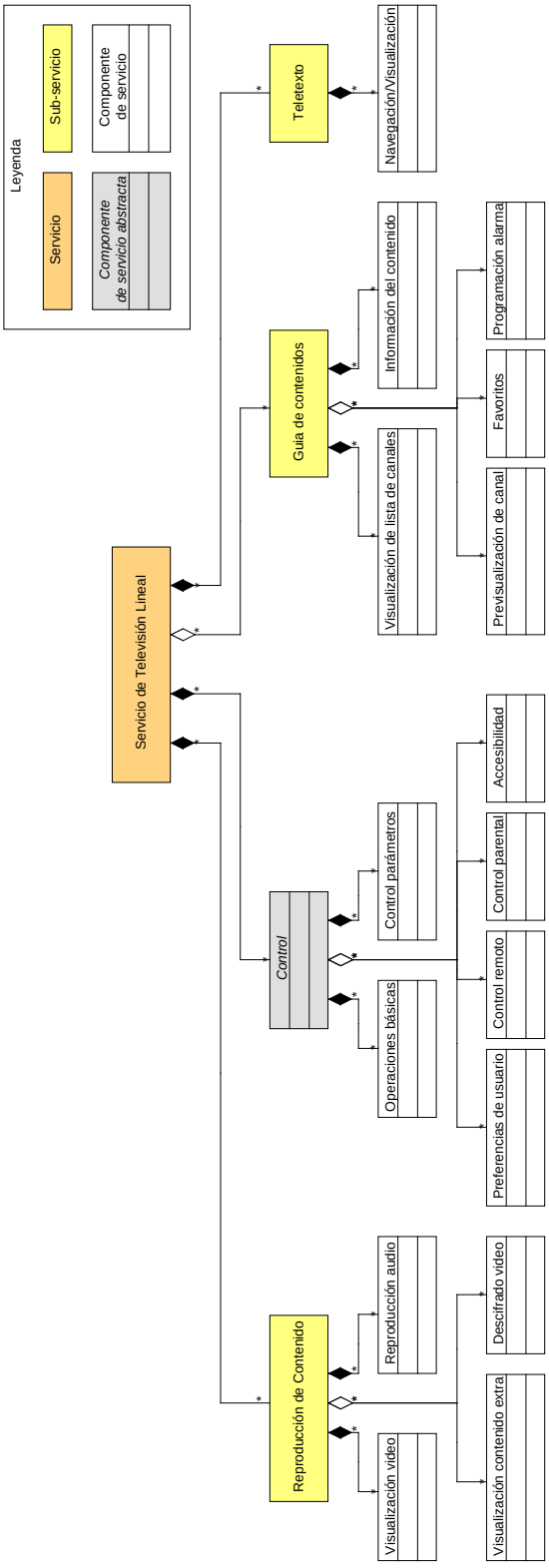


Figura A.2: Descripción del servicio de televisión lineal utilizando el modelo propuesto

El sub-servicio de Reproducción del contenido deberá contar con componentes de servicio como pueden ser: Visualización de vídeo, Reproducción de audio, Descrifrado de contenidos, DRM (opcional) y Visualización de contenido extra (opcional).

El sub-servicio de Guía de contenidos deberá estar compuesto al menos por los siguientes componentes de servicio: Visualización de la lista de canales, Visualización de información del contenido y otros componentes opcionales, como pueden ser Gestión de canales favoritos, Programación de alarmas, etc.

Por su parte, el sub-servicio de Teletexto debe contar al menos con un componente de servicio de Navegación/Visualización que permita acceder y representar la información.

Además, existen una serie de componentes de servicio que se pueden agrupar dentro de una categoría que se ha denominado "control". En esta categoría se incluyen componentes de servicio que ofrecen funcionalidades básicas de control, tanto del servicio, como del terminal que lo soporta. A esta categoría pertenecerían componentes de servicio como por ejemplo: control del terminal (operaciones básicas y control de parámetros), gestión de preferencias de usuario, control remoto del terminal, etc. Como se puede ver en la figura A.2, estos componentes se han agrupado mediante la utilización de una clase abstracta denominada "Control".

Por último, para completar esta fase de la metodología, habría que definir los atributos y métodos de cada uno de los componentes de servicio. A modo de ejemplo, en la figura A.3 se concretan los atributos y los métodos del componente de servicio "Visualización de vídeo". Como se puede ver, los atributos y métodos están expresados desde un punto de vista de alto nivel (cercano a la percepción del usuario), por lo que en el caso concreto de la visualización de vídeo, los atributos relevantes para el usuario están relacionados con la calidad del mismo (independientemente de la implementación usada para conseguir dicha calidad) y los métodos están relacionados con las operaciones que el usuario puede llevar a cabo en la reproducción del vídeo, es decir, controlar dicha reproducción (iniciar, detener, pausar, etc.).

Paso 3: definición de los bloques arquitecturales Es importante destacar que, dentro de los diferentes niveles de abstracción contemplados en el modelo, los bloques arquitecturales conforman el primer nivel que introduce aspectos propios de la estructura y la organización que debe seguir el sistema concreto que dé soporte al servicio.

Aunque en la figura A.2 no se han incluido los bloques arquitecturales (por restricciones en cuanto al tamaño de la misma), a modo de ejemplo se van a describir en detalle los bloques arquitecturales del componente de servicio "Visualización de vídeo", suponiendo que el sistema que va a dar soporte al servicio es un sistema de distribución de vídeo OTT, ver figura A.3. El primer bloque arquitectural (Servidor de vídeo) es de tipo "Contenido" y representa un repositorio donde se almacena el contenido a repro-

ducir. El segundo bloque (Canal de comunicación), de tipo Red, representa el medio y los protocolos necesarios para acceder al contenido almacenado en el Servidor de vídeo. Por último, como su nombre indica, el bloque "Player de vídeo", de tipo Equipamiento de Usuario, es el encargo de reproducir el vídeo en el dispositivo del cliente.

Paso 4: definición de la implementación El último paso en la aplicación de la metodología consiste en concretar la implementación de cada bloque arquitectural.

Siguiendo el ejemplo de los pasos anteriores, en la figura A.3 se propone una posible implementación para los bloques arquitecturales del componente de servicio "Visualización de vídeo". Como se puede ver, al ser un sistema OTT se ha utilizado una solución basada en MPEG-DASH, por lo que además del propio protocolo y del canal TCP/IP consta de un servidor HTTP y un reproductor de vídeo HTML5 (se ha seleccionado Apache y Video.js a modo de ejemplo).

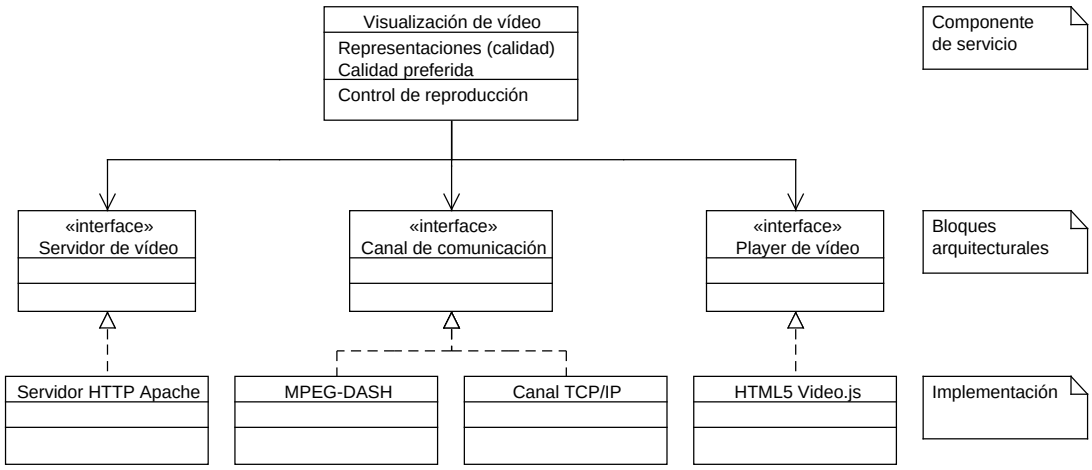


Figura A.3: Componente de servicio "Visualización de video": bloques arquitecturales e implementaciones para un sistema de vídeo OTT

A.4.1.2. Aplicación de la metodología al servicio de vídeo bajo demanda

El objetivo de esta sección no es simplemente proporcionar otro ejemplo para ilustrar la aplicación de la metodología a otro servicio, sino remarcar el carácter reutilizable de los distintos componentes del modelo.

El servicio de vídeo bajo demanda es un servicio multimedia donde cada usuario de forma individual puede seleccionar el contenido que desea consumir de una lista de contenido disponible.

En la figura A.4 se muestra la representación gráfica del servicio de VoD utilizando el modelo propuesto.

Si se compara la figura A.4 con la figura A.2 se pueden ver diversos elementos comunes. Desde el punto de vista del usuario, el servicio de televisión lineal y el servicio de vídeo bajo demanda comparten muchas características, por lo que así se representa en el modelo. Más concretamente, los sub-servicios de "Reproducción de contenido", "Control" y "Guía de contenidos" aparecen en ambos servicios. Se debe destacar también que el modelo permite la extensión de algunos de estos sub-servicios, por ejemplo, añadiendo un nuevo componente de servicio (Recomendación de contenido) al módulo de "Guía de contenidos".

Además, el servicio de VoD consta de ciertos sub-servicios como "Personal Video Recorder (PVR)", asociado a las funciones de grabación personal de contenidos; "Comunicaciones Auxiliares", que engloba servicios secundarios como integración con redes sociales, mensajería entre usuarios del servicio, etc. Por último, el sub-servicio de "Interactividad" engloba aquellos componentes de servicio que permiten a los usuarios interactuar directamente con los contenidos emitidos (compra de productos, etc.).

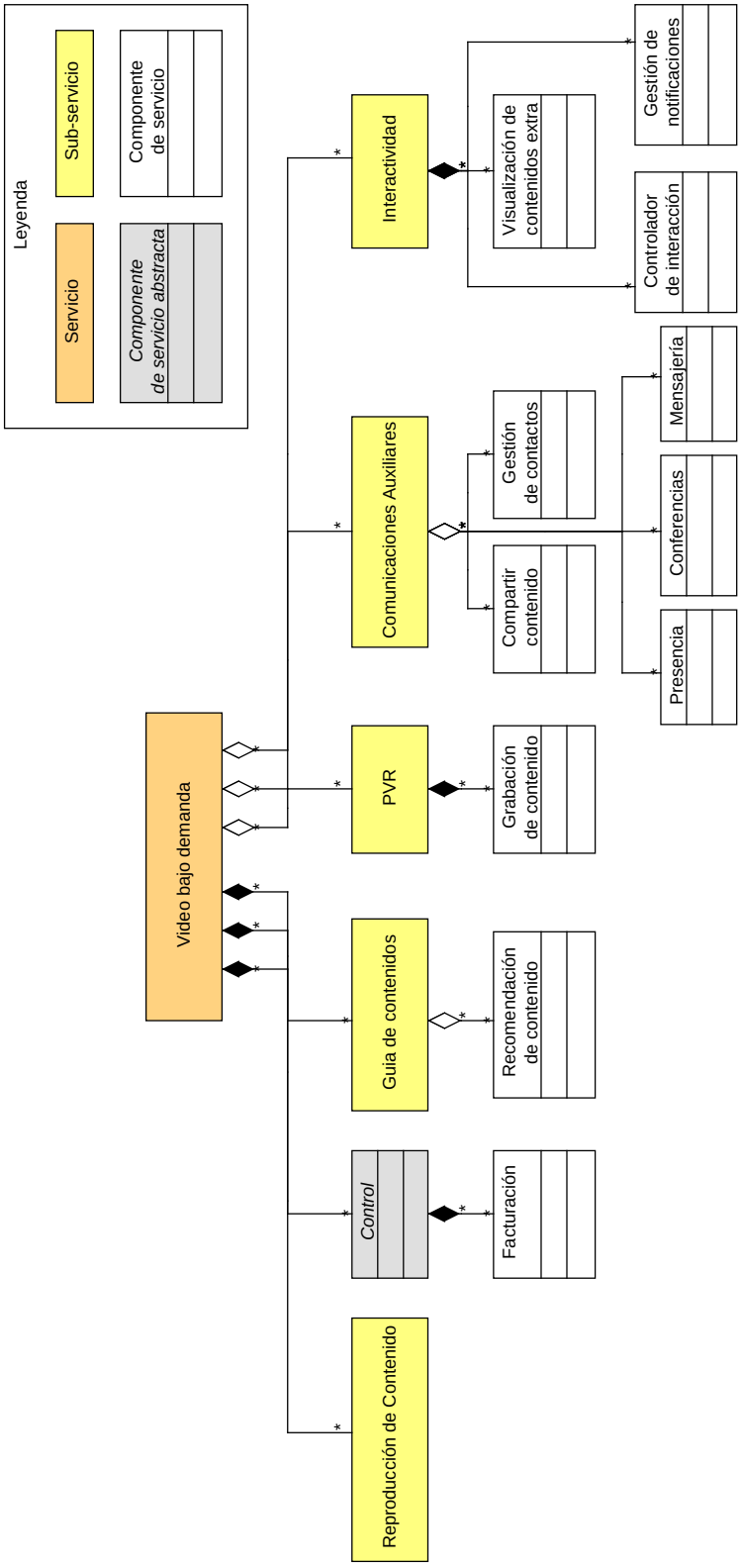


Figura A.4: Descripción del servicio de VoD utilizando el modelo propuesto

Apéndice B

Secuencias de vídeo utilizadas

En este apéndice se presenta una trama representativa de cada una de las secuencias de vídeo utilizadas en el entrenamiento y en desarrollo de los distintos modelos de calidad propuestos en esta tesis.

B.1. Modelo de calidad de vídeo

Tabla B.1: Secuencias de vídeo VQEGHD1

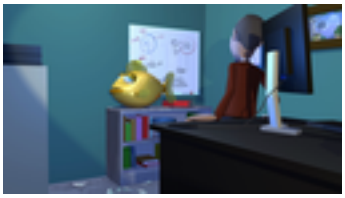
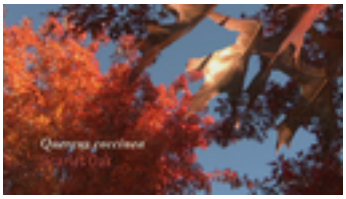
hd1src1	hd1src2	hd1src3
		
hd1src4	**hd1src5**	**hd1src6**
		
hd1src7	**hd1src8**	**hd1src9**
		

Tabla B.2: Secuencias de vídeo VQEGHD2

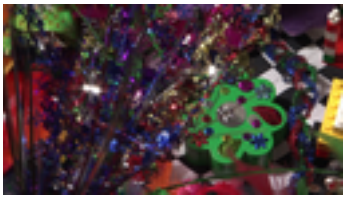
hd2src1	hd2src2	hd2src4
		
hd2src5	**hd2src6**	**hd2src7**
		
hd2src8	**hd2src9**	
		

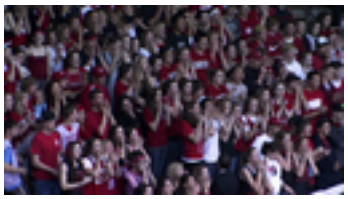
Tabla B.3: Secuencias de vídeo VQEGHD3

hd3src1	hd3src2	hd3src3
		
hd3src4	**hd3src5**	**hd3src6**
		
hd3src7	**hd3src8**	**hd3src9**
		

Tabla B.4: Secuencias de vídeo VQEGHD5

hd5src1	hd5src2	hd5src4
		
hd5src5	**hd5src6**	**hd5src8**
		
hd5src9		
		

Tabla B.5: Secuencias de vídeo VQEGHDCommonSet

cssrc11	cssrc12	cssrc13
		
cssrc14		
		

B.2. Degradación asociada al tiempo de buffering inicial

Tabla B.6: Secuencias de vídeo utilizadas en el experimento de evaluación de calidad del tiempo de buffering inicial

Game of Thrones Soundtrack v1	Nasa asteroid	DJI Phantom v1
		
Big Buck Bunny v1		
		

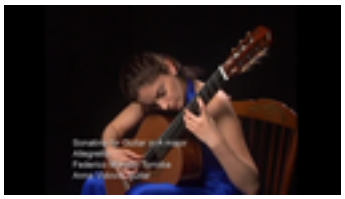
B.3. Degradación asociada al tiempo de rebuffering

Tabla B.7: Secuencias de vídeo utilizadas en el experimento de evaluación de calidad del tiempo de rebuffering

Red Bull v1	Sintel v1	Skyrim v1
		
Space to ground v1		
		

B.4. Degradación asociada al número de eventos de rebuffering

Tabla B.8: Secuencias de vídeo utilizadas en el experimento de evaluación de calidad del número de eventos de rebuffering

Ana Vidovic	Federer vs Nadal	Nasa Mars
		
Portal 2		
		

B.5. Degradación asociada a los mecanismos de adaptación de calidad

Tabla B.9: Secuencias de vídeo utilizadas en el experimento de evaluación de calidad de la adaptación de vídeo (1 de 2)

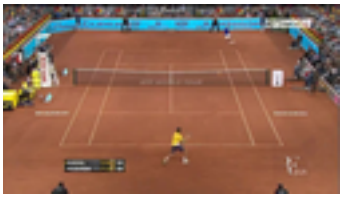
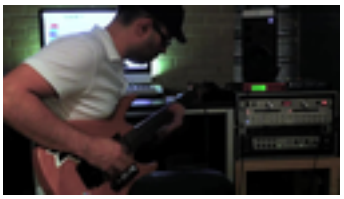
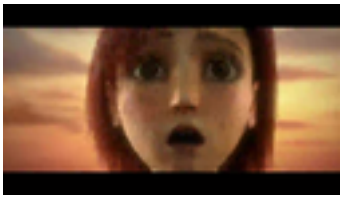
Bike v1	Bike v2	Federer v1
		
Federer v3	Game of Thrones Soundtrack v2	Madrid Aerial v1
		

Tabla B.10: Secuencias de vídeo utilizadas en el experimento de evaluación de calidad de la adaptación de vídeo (2 de 2)

Madrid Aerial v2	Madrid Aerial v3	Nasa Artic v1
		
Nasa Artic v2	Novedades F1 v1	Novedades F1 v2
		
DJI Phantom	Sintel v1	Sintel v2
		
Skyrim v1	Skyrim v2	Space to ground v1
		
Space to ground v2	Big Buck Bunny	
		

Apéndice C

Plataforma web de evaluación subjetiva de calidad de vídeo

C.1. Introducción

Como se comentó en la sección 5.3.2, en esta tesis se ha llevado a cabo un conjunto de experimentos de evaluación subjetiva de calidad de vídeo mediante la utilización de una plataforma web diseñada para ello.

El paradigma de la evaluación subjetiva de calidad basada en crowdsourcing está cobrando especial relevancia en los últimos años, por lo que han aparecido en escena diversas plataformas web que permiten llevar a cabo este tipo de experimentos de evaluación. De entre la diversidad de trabajos relacionados con la evaluación de calidad percibida mediante crowdsourcing, algunos destacan por ofrecer sus plataformas web a través de licencias de código abierto. Algunas de las plataformas disponibles más destacadas son las siguientes:

- QualityCrowd y QualityCrowd2 (desarrollado por Technische Universität München [Keimel et al., 2012]
- Web-based Subjective Quality Evaluation Platform (desarrollada por ITEC) [Rainer et al., 2013]
- Web-Enabled Subjective Test (desarrollada por NTIA) [NTIA, 2014].

El enfoque que se ha seguido en esta tesis para el desarrollo de la plataforma web de evaluación se basa en seleccionar una de las plataformas disponibles y adaptarla a las necesidades de la tesis. Tras analizar cada una de las plataformas anteriores, se ha decidido utilizar QualityCrowd2 como plataforma "base", ya que de las plataformas analizadas es la que se encuentra en una fase de desarrollo más madura. En cuanto a las plataformas descartadas, cabe destacar que WEST todavía está en fases iniciales

de desarrollo, por lo que adaptarla a las necesidades de la tesis hubiese supuesto la realización de un amplio desarrollo. En el caso de la plataforma de ITEC, aunque el código fuente está disponible, no ponen a disposición de los usuarios la estructura que debe seguir la base de datos que utiliza, por lo que se descartó su utilización.

C.2. QualityCrowd2

QualityCrowd2 es una plataforma web de evaluación de calidad de vídeo e imagen, desarrollada por la universidad de Munich. Esta plataforma está escrita en PHP y permite la definición de experimentos (o batches, según la nomenclatura de QualityCrowd2) mediante ficheros de texto (QC-scripts) que siguen una sintaxis especial definida a tal efecto. Qualitycrowd2 no requiere el uso de base de datos (los resultados se almacenan en ficheros en disco) y aunque la versión anterior soportaba la integración con sistemas como Crowdflower, Turk, etc., en la versión actual los conectores encargados de esta integración han sido eliminados a cambio de un sistema de tokens e identificadores de "workers" (usuarios encargados de realizar las evaluaciones).

QualityCrowd2 soporta diferentes tipos de respuestas: respuestas textuales libres, respuestas predefinidas en las que el usuario debe seleccionar una opción y respuestas "continuas", las cuales se implementan mediante una barra con un control deslizante, el cual puede ser situado por el usuario en la posición que desee.

En cuanto a la reproducción de contenido, el reproductor de vídeo por defecto de QualityCrowd2 es un player de vídeo Flash (qcplayer). Además, implementa un fallback a vídeo HTML5 en caso de que el cliente no disponga del plugin de Adobe Flash en el navegador.

Esta plataforma dispone de un panel de administración desde el que se gestionan los batches (creación y edición) y se visualizan los resultados. Los resultados de cada batch pueden ser exportados por la herramienta a formatos como CSV o XLSX para su posterior análisis.

C.3. Modificaciones realizadas a QualityCrowd2

Aunque Qualitycrowd2 es una plataforma bastante completa y funcional, para los intereses de la tesis se han llevado a cabo algunas modificaciones con el objetivo de adaptarla a nuestras necesidades.

C.3.1. Sustitución del reproductor de vídeo

En primer lugar, las pruebas realizadas con el reproductor de vídeo por defecto que incluye QualityCrowd2 no dieron el resultado esperado, mostrando una inestabilidad

inaceptable para la realización de las pruebas de vídeo con contenido de alta resolución. Esta inestabilidad motivó que se sustituyese el player Flash por un player de vídeo HTML5.

El player de vídeo utilizado fue Video.js [Brightcove and Zencoder, 2014]. Video.js es una librería Javascript y CSS que implementa un conjunto de controles sobre el elemento video de HTML5 con el objetivo de proporcionar un aspecto consistente entre browsers, resolviendo inconsistencias o errores y añadiendo funcionalidades no soportadas en todos los navegadores. Además, y de especial relevancia como se verá a continuación, Video.js proporciona una API en Javascript con la que poder controlar aspectos relacionados con la apariencia del player y la reproducción del contenido. La integración del player Video.js en QualityCrowd2 es relativamente sencilla y se basa en la creación de un nuevo fichero template (.tpl), que se encargue de inicializar dicho player, sustituyendo al template del player Flash por defecto.

C.3.2. Simulación de eventos de buffering inicial y rebuffering

En segundo lugar, uno de los objetivos de la tesis es estudiar el efecto que los eventos de (re)buffering tienen en la calidad percibida. Para llevar a cabo experimentos de evaluación subjetiva que estudien este tipo de degradación es necesario desarrollar un entorno que permita "reproducir" o "simular" eventos de buffering inicial y de rebuffering de manera controlada.

La primera idea que se barajó fue la de "simular" los eventos de rebuffering como parte del contenido del vídeo. Es decir, la idea consistía en introducir fragmentos de vídeo que visualmente fueran similares al comportamiento típico de los reproductores de vídeo cuando están en modo rebuffering (rueda que gira, etc.). Sin embargo, esta idea se descartó, ya que aunque técnicamente es posible, no es una solución demasiado escalable. Por ejemplo, si se quiere evaluar la calidad de una misma secuencia de vídeo cuando ésta sufre un evento de rebuffering de 3, 5 y 10 segundos, esto obligaría a crear 3 secuencias de vídeo de prueba distintas.

Tras analizar el problema, se decidió optar por otra solución mucho más escalable, sencilla y elegante que la anterior. Cuando se produce un evento de rebuffering, el reproductor de vídeo no tiene contenido para reproducir e inicia una animación para hacérselo saber al usuario. Teniendo esto en cuenta, si de alguna manera se pudiera "forzar" al reproductor de vídeo para que reproduzca dicha animación de manera controlada, la simulación del rebuffering sería totalmente creíble y no conllevaría ningún proceso de codificación de vídeo adicional. Además, si se desea evaluar la calidad de una secuencia de vídeo con distintas degradaciones, se pueden aprovechar mecanismos de caché, ya que realmente la secuencia de vídeo que reproduce el cliente sería siempre la misma, solo cambiarían las degradaciones que se realizan (en local) forzando al

reproductor de vídeo que muestre la animación de rebuffering.

Como se comentó anteriormente, una de las ventajas del player Video.js es que proporciona una API en Javascript con la que se pueden controlar algunos parámetros de la reproducción. Aprovechando esta API, se ha implementado un mecanismo que permite simular eventos de rebuffering de una duración determinada en cualquier instante del vídeo. En concreto, mediante la generación forzada del evento "waiting" se puede conseguir que el player simule un rebuffering. La generación de este evento junto con el procesado del evento "timeupdate", el cual se genera periódicamente para informar del instante de reproducción actual, permiten una simulación controlada y "on the fly" de eventos de rebuffering.

C.3.3. Extensión de la sintaxis QC-script

La última modificación que se realizó sobre la plataforma QualityCrowd2 fue la extensión de su sintaxis de definición de batches. Dicha extensión permite especificar de una manera sencilla los eventos de buffering inicial y de rebuffering que se quieren simular. Por ejemplo, si se incluye la siguiente línea, el player de vídeo HTML5 forzará 3 eventos de rebuffering en los instantes 20, 60 y 120 segundos, de duración 4, 10 y 5 segundos respectivamente.

set rebufferingsimulation "20: 4, 60: 10, 120: 5"

Una vez que se utiliza la sentencia "set rebufferingsimulation" ésta aplica a todos los vídeos que aparezcan definidos a continuación de la misma. Para definir diferentes simulaciones de rebufferings basta con volver a utilizar dicha sentencia. Si no se quieren aplicar más eventos de rebuffering se puede utilizar la siguiente sentencia:

unset rebufferingsimulation

Apéndice D

Comparativa y selección de herramientas de simulación de redes

Para llevar a cabo la simulación del servicio de distribución de vídeo mediante streaming adaptativo se ha realizado un estudio en el que se han evaluado varias herramientas de simulación de redes. El objetivo de este apéndice es analizar, comparar y seleccionar la herramienta de simulación que mejor se adapte a las necesidades de la tesis.

Actualmente existe una amplia gama de herramientas de simulación de red entre las que destacan OPNET Modeler, NS-2, NS-3, OMNeT++ y NetSim. A continuación se describen las características más destacadas de estos simuladores.

D.1. OPNET Modeler

OPNET Modeler es una de las herramientas de simulación más populares del momento, tanto en el ámbito académico como en el empresarial. Esta herramienta permite analizar distintos tipos de redes, dispositivos y aplicaciones gracias a las librerías de protocolos y tecnologías que incluye. OPNET permite al usuario realizar tres funciones principales: modelar, simular y analizar. Para las tareas de modelado, OPNET ofrece una interfaz gráfica de usuario con la que construir los escenarios a simular. La correspondencia entre los elementos de la GUI y la implementación real en el simulador se realiza mediante programación orientada a objetos en lenguaje C++. En cuanto a la simulación, OPNET soporta 4 tecnologías o métodos de simulación:

- Simulación de eventos discretos: se utilizan modelos muy detalladas que simulan explícitamente el intercambio de paquetes y mensajes. Ofrece resultados muy

fiables aunque como contrapartida los tiempos de simulación son mayores que en los otros métodos.

- Análisis de flujo: se utilizan técnicas analíticas y algoritmos para modelar el comportamiento de la red en estado estacionario. Se suele utilizar para estudiar el encaminamiento y la disponibilidad a lo largo de la red en estado estacionario. Los tiempos de simulación suelen ser más rápidos que con simulación de eventos discretos.
- ACE (Application Characterization Environment) QuickPredict: se utiliza una técnica analítica para estudiar el impacto de los parámetros de red en el tiempo de respuesta de una aplicación.
- Simulación híbrida: combina dos técnicas de simulación (analítica y discreta) para proporcionar resultados precisos y detallados para un conjunto de flujos seleccionados. Se distingue entre el tráfico de fondo (utilizado para representar la carga habitual de la red) y los flujos de aplicación que se representan con detalle utilizando modelos explícitos de tráfico. Los tiempos de simulación suelen ser más rápidos que con simulación de eventos discretos.

Con respecto al análisis, OPNET ofrece herramientas como de generación de gráficos, esquemas, estadísticas, animaciones, etc., con las que presentar los resultados de forma adecuada.

D.2. NS-2

NS-2 (Network Simulator-2) es una de las herramientas de simulación de redes de código abierto más populares, estando su uso ampliamente extendido en investigaciones académicas. La arquitectura de NS-2 está basada en C++ y OTcl (Object-oriented Tool Command Language). C++ se utiliza para definir los mecanismos internos de los objetos simulados, mientras que OTcl se utiliza para definir escenarios y topologías ensamblando y configurando los objetos involucrados. OTcl se utiliza también para programar eventos discretos a lo largo de la simulación. Existen herramientas externas que ayudan en la visualización e interpretación de trazas y resultados. Una de las más destacadas es Nam (Network Animator).

D.3. NS-3

Análogamente a NS-2, NS-3 es una herramienta de simulación de redes de código abierto (licencia GNU GPLv2) orientada a uso educacional y de investigación. NS-3 está llamado a reemplazar a NS-2, pero se debe destacar que NS-3 no es una actualización

de NS-2, sino que ha sido reescrito por completo y no es compatible hacia atrás con NS-2. Las principales diferencias con respecto a NS-2 son las siguientes:

- Núcleo escrito en C++ y Python como lenguaje de scripting.
- Mayor realismo en los elementos: diseños más cercanos a las arquitecturas reales.
- Diseño modular: permite reutilizar módulos software y reduce la necesidad de reescribir modelos.
- Soporte a virtualización.
- Framework de trazas: NS-3 permite obtener estadísticas y personalizar los resultados sin tener que reescribir el núcleo del simulador.

D.4. OMNeT++

OMNeT++ es un entorno de simulación de eventos discretos de código abierto (tiene su propia licencia) modular y de arquitectura abierta utilizado en múltiples campos, entre los que destacan arquitecturas hardware, procesos de negocio y sobre todo, redes de comunicaciones. Los módulos de OMNeT++ se escriben en C++ y se ensamblan usando un lenguaje de alto nivel (NED). OMNeT++ cuenta también con interfaz gráfico de usuario para la creación de escenarios. Se debe destacar que OMNeT++ por sí mismo no proporciona componentes específicos de simulación de redes, ni de ningún área en particular. Los componentes necesarios para realizar simulaciones están contenidos en otros paquetes como INET Framework (contiene modelos de protocolos como UDP, TCP, SCTP, IP, IPv6, etc.), MiXiM (ofrece modelos detallados de propagación de onda, interferencia y consumo de potencia para redes de sensores inalámbricas, redes ad-hoc, redes vehiculares, etc.) y Castalia (simulador de dispositivos embebidos de baja potencia). El desarrollo de estos paquetes es independiente de OMNeT++ por lo que cada uno sigue su propio ciclo de desarrollo.

D.5. NetSim

NetSim es una herramienta de simulación de redes de comunicación desarrollada y comercializada por Tetcos. NetSim se organiza en componentes, los cuales encapsulan distintos protocolos y tecnologías. Existen componentes de encaminamiento IP, TCP y UDP, MANETs, Wi-Max, CDMA, entre otros. NetSim proporciona un generador de tráfico con el que modelar transmisión de voz y datos. Ofrece también un entorno de desarrollo llamado DEN (Development Environment in NetSim) con el que los usuarios pueden escribir sus propios modelos (desarrollados en lenguaje C) y enlazarlos con

el núcleo de NetSim utilizando un conjunto de librerías que ofrece el mismo. Para el análisis de resultados NetSim cuenta con un sistema de medición de rendimiento y generación de estadísticas, además de un exportador de trazas y distintas herramientas de animación. NetSim se comercializa en dos versiones distintas: una versión estándar y una versión académica (ambas de pago) de funcionalidad limitada con respecto a la versión estándar.

D.6. Selección de la herramienta de simulación

Una vez analizadas las herramientas de simulación de redes más utilizadas del momento, se ha optado por la utilización de OMNeT++. Las principales razones que han llevado a esta decisión son las siguientes:

- El carácter modular y extensible de OMNeT++ y del framework INET es la característica que más peso ha tenido a la hora de seleccionar OMNeT++ como herramienta a utilizar. En ninguna de las herramientas analizadas existen módulos con los que simular las características propias del streaming adaptativo sobre TCP, por lo que ha sido necesario desarrollar un nuevo modelo, a partir de los que ya están desarrollados en la herramienta. En este contexto el framework INET (que implementa la torre de protocolos TCP/IP sobre OMNeT++) destaca por la claridad de su diseño y por una serie de clases e interfaces bien definidas sobre las que poder implementar nuevos modelos. Más concretamente, las clases TCPSocket junto con las clases TCPGenericCliApp y TCPGenericSrvApp han permitido el desarrollo del modelo de simulación de streaming de vídeo adaptativo sobre TCP.
- Interfaz gráfica con funcionalidades de generación de gráficas, estadísticas, animaciones, etc.
- La cantidad y calidad de la documentación es aceptable.
- Herramienta gratuita para uso académico, con licencia similar a GNU-GPL .

Bibliografía

[Akhshabi et al., 2012] Akhshabi, S., Anantakrishnan, L., Begen, A. C. and Dovrolis, C. (2012). What Happens when HTTP Adaptive Streaming Players Compete for Bandwidth? In Proceedings of the 22Nd International Workshop on Network and Operating System Support for Digital Audio and Video NOSSDAV '12 pp. 9–14, ACM, New York, NY, USA.

[Akhshabi et al., 2011] Akhshabi, S., Begen, A. C. and Dovrolis, C. (2011). An Experimental Evaluation of Rate-adaptation Algorithms in Adaptive Streaming over HTTP. In Proceedings of the Second Annual ACM Conference on Multimedia Systems MMSys '11 pp. 157–168, ACM, New York, NY, USA.

[Amazon, 2014] Amazon (2014). Amazon Mechanical Turk. `https://www.mturk.com`. [Online; accessed 12-February-2014].

[ANSI, 1996] ANSI (1996). Digital Transport of One-Way Video Signals – Parameters for Objective Performance Assessment. ANSI T1.801.03–1996.

[Argyropoulos et al., 2011] Argyropoulos, S., Raake, A., Garcia, M. N. and List, P. (2011). No-reference bit stream model for video quality assessment of h.264/AVC video based on packet loss visibility. In 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) pp. 1169–1172,.

[Asghar et al., 2009] Asghar, J., Le Faucheur, F. and Hood, I. (2009). Preserving Video Quality in IPTV Networks. Broadcasting, IEEE Transactions on *55*, 386–395.

[ASQ, 2014] ASQ (2014). Glosario American Society for Quality. `http://asq.org/glossary/q.html`. [Online; accessed 1-September-2014].

[Balachandran et al., 2012] Balachandran, A., Sekar, V., Akella, A., Seshan, S., Stoica, I. and Zhang, H. (2012). A Quest for an Internet Video Quality-of-experience Metric. In Proceedings of the 11th ACM Workshop on Hot Topics in Networks HotNets-XI pp. 97–102, ACM, New York, NY, USA.

[Balachandran et al., 2013] Balachandran, A., Sekar, V., Akella, A., Seshan, S., Stoica, I. and Zhang, H. (2013). Developing a Predictive Model of Quality of Experience for Internet Video. SIGCOMM Comput. Commun. Review *43*, 339–350.

[Banodkar et al., 2008] Banodkar, D., Ramakrishnan, K., Kalyanaraman, S., Gerber, A. and Spatscheck, O. (2008). Multicast instant channel change in IPTV systems. In Communication Systems Software and Middleware and Workshops, 2008. COMSWARE 2008. 3rd International Conference on pp. 370–379,.

[Beerends and De Caluwe, 1999] Beerends, J. G. and De Caluwe, F. E. (1999). The influence of video quality on perceived audio quality and vice versa. Journal of the Audio Engineering Society *47*, 355–362.

[Bellard, 2014] Bellard, F. (2014). FFmpeg project. `http://ffmpeg.org/`. [Online; accessed 2-July-2014].

[Besson et al., 2013] Besson, A., De Simone, F. and Ebrahimi, T. (2013). Objective quality metrics for video scalability. In 2013 20th IEEE International Conference on Image Processing (ICIP) pp. 59–63,.

[Bouch and Sasse, 1999] Bouch, A. and Sasse, M. A. (1999). Network quality of service: What do users need. In Proceedings of the 4th International Distributed Conference vol. 22, pp. 21–23,.

[Brandao and Queluz, 2010] Brandao, T. and Queluz, M. (2010). No-Reference Quality Assessment of H.264/AVC Encoded Video. IEEE Transactions on Circuits and Systems for Video Technology *20*, 1437–1447.

[Brightcove and Zencoder, 2014] Brightcove and Zencoder (2014). HTML5 Video Player. `http://www.videojs.com`. [Online; accessed 15-Septembre-2014].

[Brunnstrom et al., 2009] Brunnstrom, K., Hands, D., Speranza, F. and Webster, A. (2009). VQeg validation and ITU standardization of objective perceptual video quality metrics [Standards in a Nutshell]. IEEE Signal Processing Magazine *26*, 96–101.

[Chen et al., 2010] Chen, K.-T., Chang, C.-J., Wu, C.-C., Chang, Y.-C. and Lei, C.-L. (2010). Quadrant of euphoria: a crowdsourcing platform for QoE assessment. IEEE Network *24*, 28–35.

[Chen et al., 2009] Chen, K.-T., Wu, C.-C., Chang, Y.-C. and Lei, C.-L. (2009). A Crowdsourceable QoE Evaluation Framework for Multimedia Content. In Proceedings of the 17th ACM International Conference on Multimedia MM '09 pp. 491–500, ACM, New York, NY, USA.

[Château, 1998] Château, N. (1998). Study of the Influence of Experimental Context on the Relationships Between Audio, Video and Audiovisual Subjective Qualities. ITU-T SG-12 COM 12 (CNET/France Telecom).

[Cisco, 2014] Cisco (2014). Cisco Visual Networking Index: Forecast and Methodology, 2013–2018. `http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf`. [Online; accessed 7-October-2014].

[CNMC, 2012] CNMC (2012). Informe Anual 2012. Technical report Comisión Nacional de los Mercados y la Competencia.

[Cranley et al., 2006] Cranley, N., Perry, P. and Murphy, L. (2006). User Perception of Adapting Video Quality. International Journal of Human-Computer Studies *64*, 637–647.

[Cranley et al., 2007] Cranley, N., Perry, P. and Murphy, L. (2007). Dynamic content-based adaptation of streamed multimedia. Journal of network and computer applications *30*, 983–1006.

[Cronin and Taylor, 1992] Cronin, J. J. and Taylor, S. A. (1992). Measuring service quality: a reexamination and extension. The Journal of Marketing *56*, 55–68.

[Cronin and Taylor, 1994] Cronin, J. J. and Taylor, S. A. (1994). SERVPERF versus SERVQUAL: reconciling performance based and perceptions minus expectations measurement of service quality. The Journal of Marketing *58*, 125–131.

[de la Cruz Ramos, 2012] de la Cruz Ramos, P. (2012). Contribución a los Modelos y Metodologías para la Estimación de la Calidad Percibida por los Usuarios (QoE) a partir de Parámetros de Calidad de Red/Servicio (QoS) en Servicios Convergentes Multimedia (Triple-Play). PhD thesis, Departamento de Ingeniería de Sistemas Telemáticos - E.T.S.I. Telecomunicación (UPM).

[de la Cruz Ramos et al., 2012] de la Cruz Ramos, P., Navarro Salmerón, J., Pérez Leal, R. and González Vidal, F. (2012). Estimating Perceived Video Quality from Objective Parameters in Video over IP Services. In ICDT 2012, The Seventh International Conference on Digital Telecommunications pp. 65–68,.

[De Pessemier et al., 2013] De Pessemier, T., De Moor, K., Joseph, W., De Marez, L. and Martens, L. (2013). Quantifying the Influence of Rebuffering Interruptions on the User's Quality of Experience During Mobile Video Watching. IEEE Transactions on Broadcasting *59*, 47–61.

[Drucker, 1985] Drucker, P. (1985). Innovation and entrepreneurship. Harper & Row.

[DSL, 2006] DSL (2006). Triple-play Services Quality of Experience (QoE) requirements. DSL Forum Technical Report TR-126.

[Duan et al., 2004] Duan, L.-Y., Xu, M., Tian, Q. and Xu, C.-S. (2004). Mean shift based video segment representation and applications to replay detection. In IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004 (ICASSP '04). Proceedings. vol. 5, pp. V–709–12 vol.5,.

[Eckert et al., 2013] Eckert, M., Knoll, T. and Schlegel, F. (2013). Advanced MOS calculation for network based QoE Estimation of TCP streamed Video Services. In 2013 7th International Conference on Signal Processing and Communication Systems (ICSPCS) pp. 1–9,.

[ETSI, 2010] ETSI (2010). QoS and network performance metrics and measurement methods; Part 1: General considerations. ETSI EG 202 765-1.

[Farias and Mitra, 2005] Farias, M. and Mitra, S. (2005). No-reference video quality metric based on artifact measurements. In ICIP 2005. IEEE International Conference on Image Processing vol. 3, pp. III–141–4,.

[Figuerola Salas et al., 2013] Figuerola Salas, O., Adzic, V., Shah, A. and Kalva, H. (2013). Assessing Internet Video Quality Using Crowdsourcing. In Proceedings of the 2Nd ACM International Workshop on Crowdsourcing for Multimedia CrowdMM '13 pp. 23–28, ACM, New York, NY, USA.

[Fowler and Scott, 1997] Fowler, M. and Scott, K. (1997). UML distilled: applying the standard object modeling language. Addison-Wesley Longman Ltd., Essex, UK, UK.

[Gao et al., 2006] Gao, R., Dovrolis, C. and Zegura, E. (2006). Avoiding Oscillations Due to Intelligent Route Control Systems. In INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings pp. 1–12,.

[Garcia and Raake, 2009] Garcia, M. and Raake, A. (2009). Impairment-factor-based audio-visual quality model for IPTV. In 2009. QoMEx 2009. International Workshop on Quality of Multimedia Experience pp. 1–6, IEEE.

[Garcia et al., 2011] Garcia, M., Schleicher, R. and Raake, A. (2011). Impairment-Factor-Based Audiovisual Quality Model for IPTV: Influence of Video Resolution, Degradation Type, and Content Type. EURASIP Journal on Image and Video Processing *2011*, 629284.

[Garcia et al., 2013] Garcia, M.-N., List, P., Argyropoulos, S., Lindegren, D., Pettersson, M., Feiten, B., Gustafsson, J. and Raake, A. (2013). Parametric model for

audiovisual quality assessment in IPTV: ITU-T Rec. P.1201.2. In 2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP) pp. 482–487,.

[Geman et al., 1992] Geman, S., Bienenstock, E. and Doursat, R. (1992). Neural Networks and the Bias/Variance Dilemma. Neural Computation *4*, 1–58.

[Genbeta, 2014] Genbeta (2014). ¿Qué ha sido de los grandes proyectos de vídeo bajo demanda en España? `http://www.genbeta.com/web/que-ha-sido-de-los-grandes-proyectos-de-video-bajo-demanda-en-espana`. [Online; accessed 7-October-2014].

[Ghanbari, 2003] Ghanbari, M. (2003). Standard codecs: Image compression to advanced video coding. Number 49, IET.

[3GPP, 2013] 3GPP (2013). Services and service capabilities. 3GPP TS 22.105.

[ATSC, 2003] ATSC (2003). Relative timing of sound and vision for broadcast operations. ATSC IS-191.

[EBU, 2007] EBU (2007). The relative timing of the sound and vision components of a television signal. EBU Recommendation R37.

[IETF, 1994] IETF (1994). Integrated Services in the Internet Architecture: an overview. RFC 1633.

[IETF, 1997] IETF (1997). The Use of RSVP with IETF Integrated Services. RFC 2210.

[IETF, 1998a] IETF (1998a). A Framework for QoS-based Routing in the Internet. RFC 2386.

[IETF, 1998b] IETF (1998b). An Architecture for Differentiated Services. RFC 2475.

[IETF, 2001] IETF (2001). Multiprotocol Label Switching Architecture. RFC 3031.

[IETF, 2011] IETF (2011). Unicast-Based Rapid Acquisition of Multicast RTP Sessions. RFC 6285.

[ISO, 1993] ISO (1993). Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 2: Video. ISO/IEC 11172-2:1993.

[ISO, 2004] ISO (2004). Information technology – Coding of audio-visual objects – Part 2: Visual. ISO/IEC 14496-2:2004.

[ISO, 2005a] ISO (2005a). Quality management systems - Fundamentals and vocabulary. ISO 9000.

[ISO, 2005b] ISO (2005b). Information technology – Coding of audio-visual objects – Part 12: ISO base media file format. ISO/IEC 14496-12:2005.

[ISO, 2013a] ISO (2013a). Information technology – Generic coding of moving pictures and associated audio information – Part 1: Systems. ISO/IEC 13818-1:2013.

[ISO, 2013b] ISO (2013b). Information technology – Generic coding of moving pictures and associated audio information – Part 2: Video. ISO/IEC 13818-2:2013.

[ISO, 2013c] ISO (2013c). Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High efficiency video coding. ISO/IEC 23008-2:2013.

[ISO, 2014a] ISO (2014a). Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding. ISO/IEC 14496-10:2014.

[ISO, 2014b] ISO (2014b). Information technology - Dynamic adaptive streaming over HTTP (DASH) - Part 1: Media presentation description and segment formats. ISO/IEC 23009-1:2014.

[ITU, 1989] ITU (1989). Method for the characterization of telecommunication services supported by an ISDN and network capabilities of an ISDN. ITU-T I.130.

[ITU, 1993a] ITU (1993a). Codecs for videoconferencing using primary digital group transmission. ITU-T H.120.

[ITU, 1993b] ITU (1993b). Video codec for audiovisual services at p x 64 kbit/s. ITU-T H.261.

[ITU, 1993c] ITU (1993c). Attribute technique for the characterization of telecommunication services supported by an ISDN and network capabilities of an ISDN. ITU-T I.140.

[ITU, 1997a] ITU (1997a). Relations Between Audio, Video and Audiovisual Quality. ITU-T SG12 COM12-19-E (KPN).

[ITU, 1997b] ITU (1997b). Methods for subjective determination of transmission quality. ITU-T P.800.

[ITU, 1998a] ITU (1998a). Results of an Audiovisual Desktop Video Teleconferencing Sujective Experiment. ITU-T SG12 COM12 D.038 (NTIA/ITS).

[ITU, 1998b] ITU (1998b). Relative timing of sound and vision for broadcasting. ITU-R BT.1359-1.

[ITU, 1998c] ITU (1998c). Subjective audiovisual quality assessment methods for multimedia applications. ITU-T P.911.

[ITU, 1998d] ITU (1998d). Interactive test methods for audiovisual communications. ITU-T P.920.

[ITU, 2001] ITU (2001). Communications Quality of Service: A framework and definitions. ITU-T G.1000.

[ITU, 2003] ITU (2003). Requirements for an Objective Perceptual Multimedia Quality Model. ITU-T J.148.

[ITU, 2004a] ITU (2004a). Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference. ITU-R BT.1683.

[ITU, 2004b] ITU (2004b). Quality of Service and Network Performance. Handbook. ITU-T QoS.02.

[ITU, 2004c] ITU (2004c). Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference. ITU-T J.144.

[ITU, 2005] ITU (2005). Video coding for low bit rate communication. ITU-T H.263.

[ITU, 2006] ITU (2006). Mean Opinion Score (MOS) terminology. ITU-T P.800.1.

[ITU, 2007] ITU (2007). Framework and methodologies for the determination and application of QoS parameters. ITU-T E.802.

[ITU, 2008a] ITU (2008a). Definitions of terms related to quality of service. ITU-T E.800.

[ITU, 2008b] ITU (2008b). Quality of Experience Requirements for IPTV Services. ITU-T G.1080.

[ITU, 2008c] ITU (2008c). Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference. ITU-T J.246.

[ITU, 2008d] ITU (2008d). Objective perceptual multimedia video quality measurement in the presence of a full reference. ITU-T J.247.

[ITU, 2008e] ITU (2008e). New definitions for inclusion in Recommendation ITU-T P.10/G.100. ITU-T P.10 Amendment 2.

[ITU, 2008f] ITU (2008f). Subjective Video Quality Assessment Methods for Multimedia Applications. ITU-T P.910.

[ITU, 2009] ITU (2009). Information technology - Open Distributed Processing - Reference Model: Foundations. ITU-T X.902.

[ITU, 2010a] ITU (2010a). Perceptual video quality measurement techniques for digital cable television in the presence of a reduced reference. ITU-T J.249.

[ITU, 2010b] ITU (2010b). Reference algorithm for computing peak signal to noise ratio of a processed video sequence with compensation for constant spatial shifts, constant temporal shift, and constant luminance gain and offset. ITU-T J.340.

[ITU, 2011a] ITU (2011a). End-user multimedia QoS categories. ITU-T G.1010.

[ITU, 2011b] ITU (2011b). Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference. ITU-T J.341.

[ITU, 2011c] ITU (2011c). Objective multimedia video quality measurement of HDTV for digital cable television in the presence of a reduced reference signal. ITU-T J.342.

[ITU, 2011d] ITU (2011d). Internet protocol data communication service – IP packet transfer and availability performance parameters. ITU-T Y.1540.

[ITU, 2011e] ITU (2011e). Network performance objectives for IP-based services. ITU-T Y.1541.

[ITU, 2012a] ITU (2012a). Methodology for the subjective assessment of the quality of television pictures. ITU-R BT.500-13.

[ITU, 2012b] ITU (2012b). Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios. ITU-R BT.601.

[ITU, 2012c] ITU (2012c). Opinion Model for Video-Telephony Applications. ITU-T G.1070.

[ITU, 2012d] ITU (2012d). Information technology - Generic coding of moving pictures and associated audio information: Video. ITU-T H.262.

[ITU, 2012e] ITU (2012e). Parametric non-intrusive assessment of audiovisual media streaming quality - Higher resolution application area. ITU-T P.1201.2.

[ITU, 2013] ITU (2013). High efficiency video coding. ITU-T H.265.

[ITU, 2014a] ITU (2014a). Estimating End-to-End Performance in IP Networks for Data Applications. ITU-T G.1030.

[ITU, 2014b] ITU (2014b). The E-model: a computational model for use in transmission planning. ITU-T G.107.

[ITU, 2014c] ITU (2014c). Advanced video coding for generic audiovisual services. ITU-T H.264.

[NTIA, 2011] NTIA (2011). Batch Video Quality Metric (BVQM) Software. `http://www.its.bldrdoc.gov/resources/video-quality-research/guides-and-tutorials/description-of-vqm-tools.aspx`. [Online; accessed 10-July-2014].

[NTIA, 2014] NTIA (2014). Web-Enabled Subjective Test (WEST). `http://www.its.bldrdoc.gov/resources/video-quality-research/web-enabled-subjective-test-(west).aspx`. [Online; accessed 12-August-2014].

[VQEG, 2011] VQEG (2011). HDTV Phase I Final Report. VQEG HDTV Project.

[Godana et al., 2009] Godana, B., Kooij, R. E. and Ahmed, K. (2009). Impact of advertisements during channel zapping on quality of experience. In ICNS'09. Fifth International Conference on Networking and Services pp. 249–254, IEEE.

[Gouache et al., 2011] Gouache, S., Bichot, G., Bsila, A. and Howson, C. (2011). Distributed amp; adaptive HTTP streaming. In 2011 IEEE International Conference on Multimedia and Expo (ICME) pp. 1–6,.

[Grönroos, 1984] Grönroos, C. (1984). A service quality model and its marketing implications. European Journal of marketing *18*, 36–44.

[Gustafsson et al., 2008] Gustafsson, J., Heikkila, G. and Pettersson, M. (2008). Measuring multimedia quality in mobile networks with an objective parametric model. In ICIP 2008. 15th IEEE International Conference on Image Processing pp. 405–408,.

[Hands, 2004] Hands, D. (2004). A basic multimedia quality model. IEEE Transactions on Multimedia *6*, 806–816.

[Hardy, 2001] Hardy, W. C. (2001). QoS: Measurement and Evaluation of Telecommunications Quality of Service. John Wiley & Sons, Inc., New York, NY, USA.

[Hemerotek, 2014] Hemerotek (2014). La joya de Prisa: Yomvi dispara su base de usuarios (+70 %) en 2013, hasta medio millón. http://hemerotek.com/2014/03/07/la-joya-de-prisa-yomvi-dispara-su-base-de-usuarios-70-en-2013-hasta-medio-millon/. [Online; accessed 7-October-2014].

[Hernando et al., 2013] Hernando, D., de Vergara, J., Madrigal, D. and Mata, F. (2013). Evaluating quality of experience in IPTV services using MPEG frame loss rate. In Smart Communications in Network Technologies (SaCoNeT), 2013 International Conference on vol. 03, pp. 1–5,.

[Hossfeld et al., 2012] Hossfeld, T., Egger, S., Schatz, R., Fiedler, M., Masuch, K. and Lorentzen, C. (2012). Initial delay vs. interruptions: Between the devil and the deep blue sea. In 2012 Fourth International Workshop on Quality of Multimedia Experience (QoMEX) pp. 1–6,.

[Hossfeld et al., 2011] Hossfeld, T., Seufert, M., Hirth, M., Zinner, T., Tran-Gia, P. and Schatz, R. (2011). Quantification of YouTube QoE via Crowdsourcing. In 2011 IEEE International Symposium on Multimedia (ISM) pp. 494–499,.

[Hurst et al., 2004] Hurst, W., Gotz, G. and Lauer, T. (2004). New methods for visual information seeking through video browsing. In Eighth International Conference on Information Visualisation, 2004 pp. 450–455,.

[Huynh-Thu et al., 2011] Huynh-Thu, Q., Garcia, M. N., Speranza, F., Corriveau, P. and Raake, A. (2011). Study of Rating Scales for Subjective Quality Assessment of High-Definition Video. IEEE Transactions on Broadcasting *57*, 1–14.

[Jain, 1989] Jain, A. K. (1989). Fundamentals of digital image processing. Prentice-Hall, Inc.

[Jiang et al., 2012] Jiang, J., Sekar, V. and Zhang, H. (2012). Improving Fairness, Efficiency, and Stability in HTTP-based Adaptive Video Streaming with FESTIVE. In Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies CoNEXT '12 pp. 97–108, ACM, New York, NY, USA.

[Jin et al., 2007] Jin, S. H., Kim, C. S., Seo, D. J. and Ro, Y.-M. (2007). Quality Measurement Modeling on Scalable Video Applications. In MMSP 2007. IEEE 9th Workshop on Multimedia Signal Processing. pp. 131–134,.

[Joly et al., 2001] Joly, A., Montard, N. and Buttin, M. (2001). Audio-visual quality and interactions between television audio and video. In Sixth International Symposium on Signal Processing and its Applications vol. 2, pp. 438–441 vol.2,.

[Joskowicz et al., 2009] Joskowicz, J., López-Ardao, J.-C., González Ortega, M. and García, C. (2009). A Mathematical Model for Evaluating the Perceptual Quality of Video. In Future Multimedia Networking, (Mauthe, A., Zeadally, S., Cerqueira, E. and Curado, M., eds), vol. 5630, of Lecture Notes in Computer Science pp. 164–175. Springer Berlin Heidelberg.

[Karczewicz and Kurceren, 2003] Karczewicz, M. and Kurceren, R. (2003). The SP- and SI-frames design for H.264/AVC. IEEE Transactions on Circuits and Systems for Video Technology *13*, 637–644.

[Kawano et al., 2010] Kawano, T., Yamagishi, K., Watanabe, K. and Okamoto, J. (2010). No reference video-quality-assessment model for video streaming services. In 2010 18th International Packet Video Workshop (PV) pp. 158–164,.

[Keimel et al., 2012] Keimel, C., Habigt, J., Horch, C. and Diepold, K. (2012). QualityCrowd - A framework for crowd-based quality evaluation. In Picture Coding Symposium (PCS), 2012 pp. 245–248,.

[Keimel et al., 2009] Keimel, C., Oelbaum, T. and Diepold, K. (2009). No-reference video quality evaluation for high-definition video. In ICASSP 2009. IEEE International Conference on Acoustics, Speech and Signal Processing pp. 1145–1148,.

[Kooij and Geijer, 2012] Kooij, R. E. and Geijer, M. (2012). Impact of Gaming during Channel Zapping on Quality of Experience. In ICNS 2012. The Eighth International Conference on Networking and Services pp. 144–148, IARIA.

[Kooij et al., 2006] Kooij, R. E., Kamal, A. and Brunnström, K. (2006). Perceived quality of channel zapping. In Communication Systems and Networks pp. 156–159,.

[Kooij et al., 2009a] Kooij, R. E., Klos, V., Godana, B. E., Nicolai, F. and Ahmed, K. (2009a). Optimising the Quality of Experience during Channel Zapping. International Journal On Advances in Systems and Measurements *2*, 204–213.

[Kooij et al., 2009b] Kooij, R. E., Nicolai, F., Ahmed, K. and Brunnström, K. (2009b). Model validation of channel zapping quality. In IS&T/SPIE Electronic Imaging pp. 72401R–72401R, International Society for Optics and Photonics.

[Krishnan and Sitaraman, 2012] Krishnan, S. S. and Sitaraman, R. K. (2012). Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-experimental Designs. In Proceedings of the 2012 ACM Conference on Internet Measurement Conference IMC '12 pp. 211–224, ACM, New York, NY, USA.

[Leister et al., 2011] Leister, W., Boudko, S. and Halbach Røssvoll, T. (2011). Adaptive video streaming through estimation of subjective video quality. International Journal On Advances in Systems and Measurements *4*, 109–121.

[Li et al., 2000] Li, F. C., Gupta, A., Sanocki, E., He, L.-w. and Rui, Y. (2000). Browsing Digital Video. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '00 pp. 169–176, ACM, New York, NY, USA.

[Libav, 2014] Libav (2014). Open source audio and video processing tools. `http://libav.org/`. [Online; accessed 2-July-2014].

[LRG, 2013] LRG (2013). DVRs leveling off at about half of all tv households. `http://www.leichtmanresearch.com/press/120613release.pdf`. [Online; accessed 7-October-2014].

[M2M, 2014] M2M (2014). OTT Watch: Connected Device Penetration Spikes, as Does Amazon Prime. `http://www.m2mevolution.com/topics/m2mevolution/articles/377714-ott-watch-connected-device-penetration-spikes-as-does.htm`. [Online; accessed 7-October-2014].

[Ma et al., 2012] Ma, Z., Xu, M., Ou, Y.-F. and Wang, Y. (2012). Modeling of Rate and Perceptual Quality of Compressed Video as Functions of Frame Rate and Quantization Stepsize and Its Applications. IEEE Transactions on Circuits and Systems for Video Technology *22*, 671–682.

[Maki et al., 2013] Maki, T., Kukolj, D., Dordevic, D. and Varela, M. (2013). A reduced-reference parametric model for audiovisual quality of IPTV services. In 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX) pp. 6–11,.

[Microworkers, 2014] Microworkers (2014). Microworkers.com. `https://microworkers.com`. [Online; accessed 12-February-2014].

[Mok et al., 2011] Mok, R., Chan, E. and Chang, R. (2011). Measuring the quality of experience of HTTP video streaming. In 2011 IFIP/IEEE International Symposium on Integrated Network Management (IM) pp. 485–492,.

[Mok et al., 2012] Mok, R. K. P., Luo, X., Chan, E. W. W. and Chang, R. K. C. (2012). QDASH: A QoE-aware DASH System. In Proceedings of the 3rd Multimedia Systems Conference MMSys '12 pp. 11–22, ACM, New York, NY, USA.

[Naccari et al., 2009] Naccari, M., Tagliasacchi, M. and Tubaro, S. (2009). No-reference Video Quality Monitoring for H.264/AVC Coded Video. Trans. Multi. *11*, 932–946.

[Ndjiki-Nya et al., 2003] Ndjiki-Nya, P., Makai, B., Blattermann, G., Smolic, A., Schwarz, H. and Wiegand, T. (2003). Improved H.264/AVC coding using texture analysis and synthesis. In ICIP 2003. Proceedings. 2003 International Conference on Image Processing vol. 3, pp. III–849–52 vol.2,.

[Netflix, 2008] Netflix (2008). Encoding for streaming. `http://blog.netflix.com/2008/11/encoding-for-streaming.html`. [Online; accessed 7-October-2014].

[Netflix, 2013] Netflix (2013). A Brief History of Netflix Streaming. `http://blog.streamingmedia.com/wp-content/uploads/2013/07/2013SMEast-C101.pdf`. [Online; accessed 7-October-2014].

[Netflix, 2014] Netflix (2014). Delivering Breaking Bad on Netflix in Ultra HD 4K. `http://techblog.netflix.com/2014/06/delivering-netflix-in-ultra-hd-4k.html`. [Online; accessed 7-October-2014].

[Nguyen and Zakhor, 2004] Nguyen, T. and Zakhor, A. (2004). Multiple sender distributed video streaming. IEEE Transactions on Multimedia *6*, 315–326.

[Nielsen, 1994] Nielsen, J. (1994). Usability engineering. Elsevier.

[Okamoto et al., 2009] Okamoto, J., Watanabe, K., Honda, A., Uchida, M. and Hangai, S. (2009). HDTV objective video quality assessment method applying fuzzy measure. In QoMEx 2009. International Workshop on Quality of Multimedia Experience pp. 168–173,.

[Oliver, 2009] Oliver, R. L. (2009). Satisfaction: A behavioral perspective on the consumer. Second edition, ME Sharpe.

[Open IPTV Forum, 2008a] Open IPTV Forum (2008a). Service and Platform Requirements.

[Open IPTV Forum, 2008b] Open IPTV Forum (2008b). Services and Functions for Release 2.

[Ou et al., 2011a] Ou, Y.-F., Ma, Z., Liu, T. and Wang, Y. (2011a). Perceptual Quality Assessment of Video Considering Both Frame Rate and Quantization Artifacts. IEEE Transactions on Circuits and Systems for Video Technology *21*, 286–298.

[Ou et al., 2011b] Ou, Y.-F., Xue, Y., Ma, Z. and Wang, Y. (2011b). A perceptual video quality model for mobile platform considering impact of spatial, temporal, and amplitude resolutions. In 2011 IEEE 10th IVMSP Workshop pp. 117–122,.

[Ovum, 2014] Ovum (2014). Spain TV update, August 2014: Battle of the telcos, as Telefonica and Vodafone make key M&A plays.

[Oyman and Singh, 2012] Oyman, O. and Singh, S. (2012). Quality of experience for HTTP adaptive streaming services. IEEE Communications Magazine *50*, 20–27.

[Padhye et al., 2000] Padhye, J., Firoiu, V., Towsley, D. F. and Kurose, J. F. (2000). Modeling TCP Reno performance: a simple model and its empirical validation. IEEE/ACM Trans. Netw. *8*, 133–145.

[Parasuraman et al., 1988] Parasuraman, A., Zeithaml, V. and Berry, L. (1988). SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing *64*, 12–40.

[Parasuraman et al., 1991] Parasuraman, A., Zeithaml, V. and Berry, L. (1991). Refinement and reassessment of the SERVQUAL scale. Journal of Retailing *67*, 420–450.

[Pastrana-Vidal et al., 2003] Pastrana-Vidal, R., Colomes, C. Gicquel, J. and Cherifi, H. (2003). Caractérisation Perceptuelle des Interactions Audiovisuelles: Revue. In Proc. of CORESA-2003 Conference en Compresion et Representation des Signaux Audiovisuels.

[Patrick Le Callet and Perkis, 2013] Patrick Le Callet, S. M. and Perkis, A. (2013). Qualinet White Paper on Definitions of Quality of Experience.

[Pérez et al., 2011] Pérez, P., Gutierrez, J., Ruiz, J. and Garcia, N. (2011). Qualitative Monitoring of Video Quality of Experience. In 2011 IEEE International Symposium on Multimedia (ISM) pp. 470–475,.

[Pham, 2012] Pham, T. (2012). Image texture analysis using geostatistical information entropy. In 2012 6th IEEE International Conference Intelligent Systems (IS) pp. 353–356,.

[Pinson and Wolf, 2003] Pinson, M. and Wolf, S. (2003). Comparing subjective video quality testing methodologies. In SPIE Video Communications and Image Processing Conference pp. 8–11,.

[Pinson and Wolf, 2004] Pinson, M. and Wolf, S. (2004). A new standardized method for objectively measuring video quality. IEEE Transactions on Broadcasting *50*, 312–322.

[PwC, 2013] PwC (2013). Consumer Intelligence Series: Video content consumption. `http://www.pwc.com/us/en/industry/entertainment-media/publications/consumer-intelligence-series/assets/pwc-consumer-intelligence-series-product-services-innovation.pdf`. [Online; accessed 7-October-2014].

[Rainer et al., 2013] Rainer, B., Waltl, M. and Timmerer, C. (2013). A web based subjective evaluation platform. In 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX) pp. 24–25,.

[Ramos et al., 2011] Ramos, F. M., Crowcroft, J., Gibbens, R. J., Rodriguez, P. and White, I. H. (2011). Reducing channel change delay in IPTV by predictive pre-joining of TV channels. Signal Processing: Image Communication *26*, 400 – 412.

[Ries et al., 2007] Ries, M., Crespi, C., Nemethova, O. and Rupp, M. (2007). Content Based Video Quality Estimation for H.264/AVC Video Streaming. In WCNC 2007. IEEEWireless Communications and Networking Conference pp. 2668–2673,.

[Saad and Bovik, 2012] Saad, M. A. and Bovik, A. C. (2012). Blind quality assessment of videos using a model of natural scene statistics and motion coherency. In Asilomar Conference on Signals, Systems, and Computers pp. 332–336,.

[Seshadrinathan and Bovik, 2010] Seshadrinathan, K. and Bovik, A. (2010). Motion Tuned Spatio-Temporal Quality Assessment of Natural Videos. IEEE Transactions on Image Processing *19*, 335–350.

[Setton and Girod, 2005] Setton, E. and Girod, B. (2005). Video streaming with SP and SI frames. In Visual Communications and Image Processing pp. 59606F–59606F, International Society for Optics and Photonics.

[Shorten et al., 2006] Shorten, R., Wirth, F. and Leith, D. (2006). A positive systems model of TCP-like congestion control: asymptotic results. IEEE/ACM Transactions on Networking *14*, 616–629.

[Siebert et al., 2009] Siebert, P., Van Caenegem, T. and Wagner, M. (2009). Analysis and Improvements of Zapping Times in IPTV Systems. Broadcasting, IEEE Transactions on *55*, 407–418.

[Singh et al., 2012] Singh, K., Hadjadj-Aoul, Y. and Rubino, G. (2012). Quality of experience estimation for adaptive HTTP/TCP video streaming using H.264/AVC. In 2012 IEEE Consumer Communications and Networking Conference (CCNC) pp. 127–131,.

[Tan et al., 2006] Tan, X., Gustafsson, J. and Heikkilä, G. (2006). Perceived video streaming quality under initial buffering and rebuffering degradations. In MESAQIN Conference (June 2006) vol. 90,.

[Tse et al., 1999] Tse, T., Vegh, S., Shneiderman, B. and Marchionini, G. (1999). An Exploratory Study of Video Browsing, User Interface Designs and Research Methodologies: Effectiveness in Information Seeking Tasks. In Proceedings of the Annual

Meeting-American Society For Information Science vol. 36, pp. 681–692, Information Today; 1998.

[Van Wallendael et al., 2012] Van Wallendael, G., Van Lancker, W., De Cock, J., Lambert, P., Macq, J.-F. and Van De Walle, R. (2012). Fast Channel Switching Based on SVC in IPTV Environments. Broadcasting, IEEE Transactions on *58*, 57–65.

[Wang et al., 2008] Wang, B., Kurose, J., Shenoy, P. and Towsley, D. (2008). Multimedia Streaming via TCP: An Analytic Performance Study. ACM Trans. Multimedia Comput. Commun. Appl. *4*, 16:1–16:22.

[Wang et al., 2004] Wang, Z., Lu, L. and Bovik, A. C. (2004). Video quality assessment based on structural distortion measurement. Signal Processing: Image Communication *19*, 121 – 132.

[Webster et al., 1993] Webster, A. A., Jones, C. T., Pinson, M. H., Voran, S. D. and Wolf, S. (1993). An Objective Video Quality Assessment System Based on Human Perception. In in SPIE Human Vision, Visual Processing, and Digital Display IV pp. 15–26,.

[Wiegand et al., 2003] Wiegand, T., Sullivan, G., Bjontegaard, G. and Luthra, A. (2003). Overview of the H.264/AVC video coding standard. IEEE Transactions on Circuits and Systems for Video Technology *13*, 560–576.

[Winkler and Faller, 2005] Winkler, S. and Faller, C. (2005). Audiovisual quality evaluation of low-bitrate video. In Proceedings of SPIE International Symposium on Human Vision and Electronic Imaging pp. 139–148, International Society for Optics and Photonics.

[Winkler and Faller, 2006] Winkler, S. and Faller, C. (2006). Perceived Audiovisual Quality of Low-Bitrate Multimedia Content. IEEE Transactions on Multimedia *8*, 973–980.

[Winkler and Mohandas, 2008] Winkler, S. and Mohandas, P. (2008). The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics. IEEE Transactions on Broadcasting *54*, 660–668.

[Wolf and Pinson, 2007] Wolf, S. and Pinson, M. (2007). Application of the NTIA general video quality metric (VQM) to HDTV quality monitoring. In Proceedings of The Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), Scottsdale, AZ, USA.

[Wolf and Pinson, 2011] Wolf, S. and Pinson, M. (2011). Video Quality Model for Variable Frame Delay (VQM_VFD). NTIA Technical Memorandum TM-11-482.

[Wulf and Zolzer, 2013] Wulf, S. and Zolzer, U. (2013). About the imperfection of objective quality metrics on high-definition video content. In 2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP) pp. 384–389,.

[Xu et al., 2012] Xu, Q., Huang, Q. and Yao, Y. (2012). Online Crowdsourcing Subjective Image Quality Assessment. In Proceedings of the 20th ACM International Conference on Multimedia MM '12 pp. 359–368, ACM, New York, NY, USA.

[Xu et al., 2010] Xu, T., Ye, B., Wang, Q., Li, W., Lu, S. and Fu, X. (2010). APEX: A personalization framework to improve quality of experience for DVD-like functions in P2P VoD applications. In 2010 18th International Workshop on Quality of Service (IWQoS) pp. 1–9,.

[Yang et al., 2005] Yang, F., Wan, S., Chang, Y. and Wu, H. R. (2005). A novel objective no-reference metric for digital video quality assessment. IEEE Signal Processing Letters *12*, 685–688.

[Yang et al., 2007] Yang, K.-C., Guest, C., El-Maleh, K. and Das, P. (2007). Perceptual Temporal Quality Metric for Compressed Video. IEEE Transactions on Multimedia *9*, 1528–1535.

[Yang et al., 2009] Yang, X., Gjoka, M., Chhabra, P., Markopoulou, A. and Rodriguez, P. (2009). Kangaroo: Video Seeking in P2P Systems. In Proceedings of the 8th International Conference on Peer-to-peer Systems IPTPS'09 pp. 6–6, USENIX Association, Berkeley, CA, USA.

[Zencoder, 2010] Zencoder (2010). Web video stats. `http://blog.zencoder.com/2010/12/31/web-video-stats-december-2010/`. [Online; accessed 19-July-2012].

[Zink et al., 2003] Zink, M., Künzel, O., Schmitt, J. and Steinmetz, R. (2003). Subjective Impression of Variations in Layer Encoded Videos. In Proceedings of the 11th International Conference on Quality of Service IWQoS'03 pp. 137–154, Springer-Verlag, Berlin, Heidelberg.

Acrónimos

3GPP 3rd Generation Partnership Project

ACR Absolute Category Rating

ASI Average Spatial Information

ATI Average Temporal Information

ATSC Advanced Television System Committee

AVC Advanced Video Coding

CAGR Compound annual growth rate

CDN Content Delivery Network

DASH Dynamic Adaptive Streaming over HTTP

DCT Discrete Cosine Transform

DMOS Difference Mean Opinion Score

DSCQS Double Stimulus Continuous Quality Scale

EBU European Broadcasting Union

FR Full Reference

GoP Group of Pictures

HD High Definition

HDTV High Definition Television

HRC Hypothetical Reference Circuit

HTTP Hypertext Transfer Protocol

IDR Instantaneous Decoding Refresh

IETF Internet Engineering Task Force

IGMP Internet Group Management Protocol

IP Internet Protocol

IPTV Internet Protocol Television

ISO International Organization for Standardization

ITU International Telecommunication Union

MOS Mean Opinion Score

MPD Media Presentation Description

MPEG The Moving Picture Experts Group

MPEG-DASH The Moving Picture Experts Group - Dynamic Adaptive Streaming over HTTP

MSE Mean Squared Error

NAL Network Abstraction Layer

NR No Reference

NTIA National Telecommunications and Information Administration

OMG Object Management Group

OTT Over-The-Top

P2P Peer-to-peer

PCA Principal Component Analysis

PSNR Peak Signal-to-Noise Ratio

PVR Personal Video Recorder

QoE Quality of Experience

QoS Quality of Service

RDSI Red Digital de Servicios Integrados

RMSE Root Mean Squared Error

RR Reduced Reference

RTP Real-time Transport Protocol

RTT Round-Trip delay Time

SAP Stream Access Point

SD Standard Definition

SDTV Standard Definition Television

SI Spatial Information

STB Set-Top Box

SVC Scalable Video Coding

TCP Transmission Control Protocol

TI Temporal Information

UHDV Ultra High Definition Video

UML Unified Modeling Language

VoD Video on Demand

VQEG Video Quality Experts Group

VQM Video Quality Model

VQM_VFD Video Quality Model for Variable Frame Delay